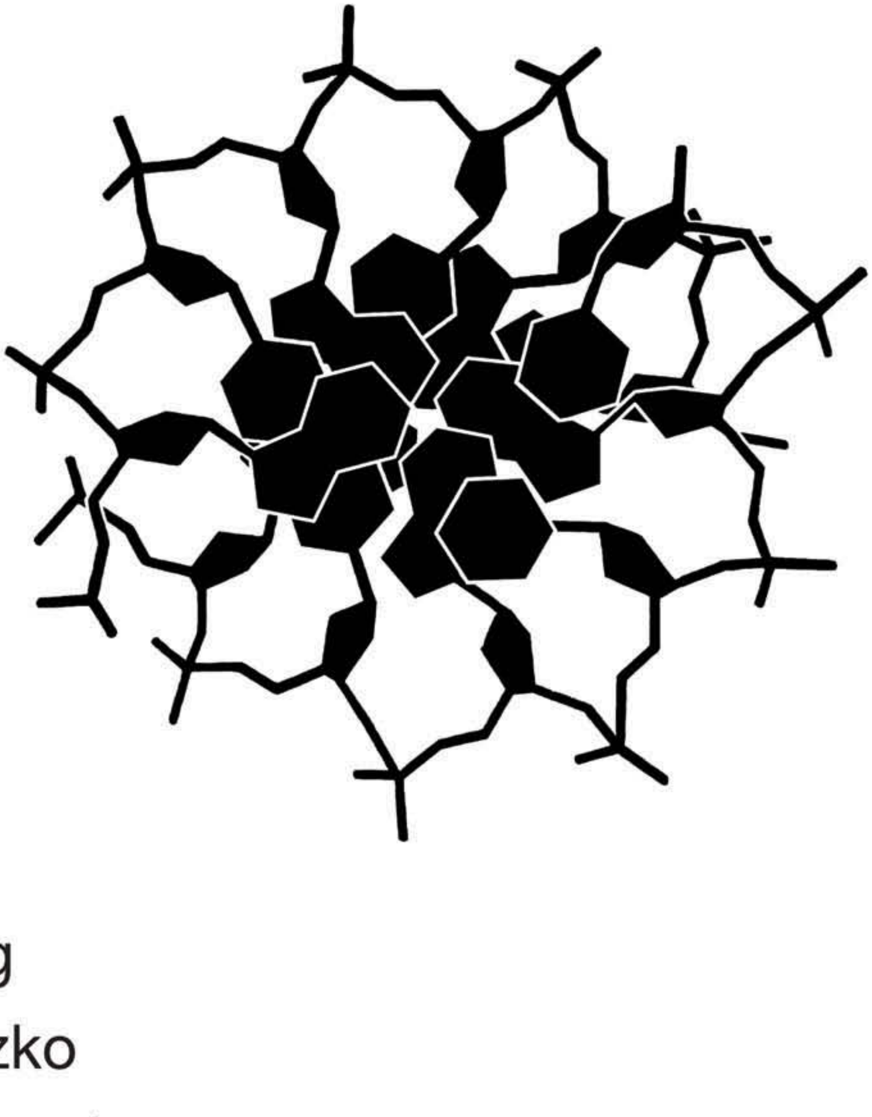
Biochemistry

EIGHTH EDITION



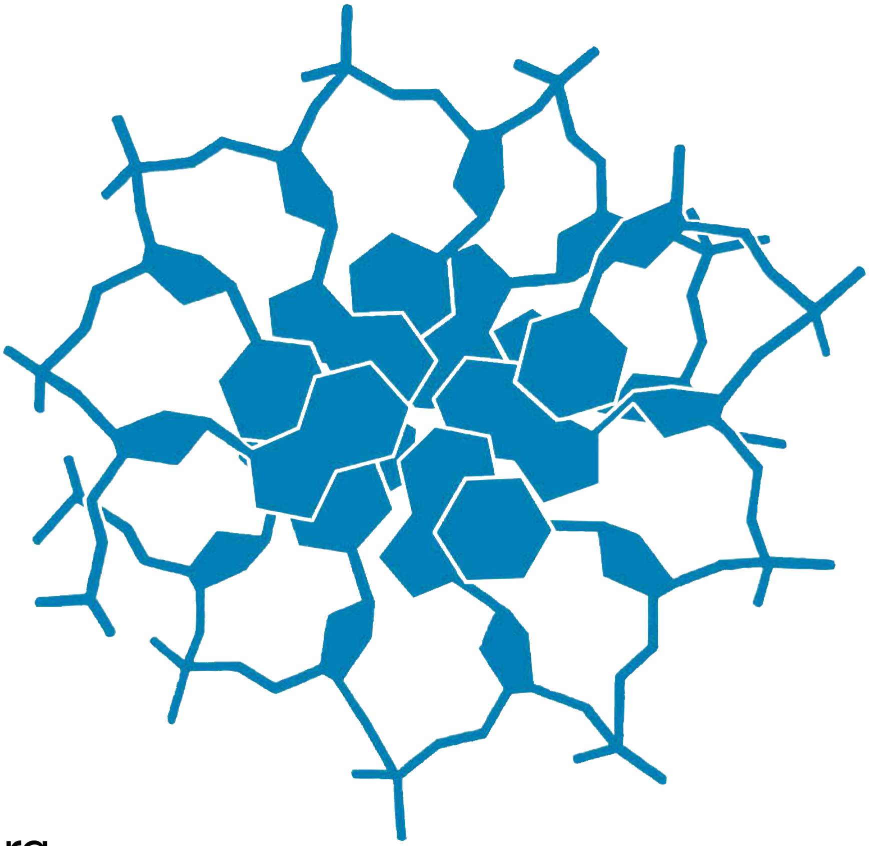
Jeremy M. Berg
John L. Tymoczko
Gregory J. Gatto, Jr.
Lubert Stryer

VISIT…

LANZAROTE
Caliente.COM

Biochemistry

EIGHTH EDITION



Jeremy M. Berg
John L. Tymoczko
Gregory J. Gatto, Jr.
Lubert Stryer

W. H. Freeman
& Company
A Macmillan Education Imprint

Publisher: Kate Ahr Parker
Senior Acquisitions Editor: Lauren Schultz
Developmental Editor: Irene Pech
Editorial Assistants: Shannon Moloney and Nandini Ahuja
Senior Project Editor: Denise Showers with Sherrill Redd
Manuscript Editors: Irene Vartanoff and Mercy Heston
Cover and Interior Design: Vicki Tomaselli
Illustrations: Jeremy Berg with Network Graphics, Gregory J. Gatto, Jr.
Illustration Coordinator: Janice Donnola
Photo Editor: Christine Buese
Photo Researcher: Jacquelyn Wong
Production Coordinator: Paul Rohloff
Executive Media Editor: Amanda Dunning
Media Editor: Donna Brodman
Executive Marketing Manager: Sandy Lindelof
Composition: Aptara®, Inc.
Printing and Binding: RR Donnelley

Library of Congress Control Number: 2014950359

Gregory J. Gatto, Jr., is an employee of GlaxoSmithKline (GSK), which has not supported or funded this work in any way. Any views expressed herein do not necessarily represent the views of GSK.

ISBN-13: 978-1-4641-2610-9
ISBN-10: 1-4641-2610-0

©2015, 2012, 2007, 2002 by W. H. Freeman and Company; © 1995, 1988, 1981, 1975 by Lubert Stryer

All rights reserved

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010

www.whfreeman.com

To our teachers and our students

ABOUT THE AUTHORS

JEREMY M. BERG received his B.S. and M.S. degrees in Chemistry from Stanford (where he did research with Keith Hodgson and Lubert Stryer) and his Ph.D. in Chemistry from Harvard with Richard Holm. He then completed a postdoctoral fellowship with Carl Pabo in Biophysics at Johns Hopkins University School of Medicine. He was an Assistant Professor in the Department of Chemistry at Johns Hopkins from 1986 to 1990. He then moved to Johns Hopkins University School of Medicine as Professor and Director of the Department of Biophysics and Biophysical Chemistry, where he remained until 2003. He then became Director of the National Institute of General Medical Sciences at the National Institutes of Health. In 2011, he moved to the University of Pittsburgh where he is now Professor of Computational and Systems Biology and Pittsburgh Foundation Professor and Director of the Institute for Personalized Medicine. He served as President of the American Society for Biochemistry and Molecular Biology from 2011–2013. He is a Fellow of the American Association for the Advancement of Science and a member of the Institute of Medicine of the National Academy of Sciences. He received the American Chemical Society Award in Pure Chemistry (1994) and the Eli Lilly Award for Fundamental Research in Biological Chemistry (1995), was named Maryland Outstanding Young Scientist of the Year (1995), received the Harrison Howe Award (1997), and received public service awards from the Biophysical Society, the American Society for Biochemistry and Molecular Biology, the American Chemical Society, and the American Society for Cell Biology. He also received numerous teaching awards, including the W. Barry Wood Teaching Award (selected by medical students), the Graduate Student Teaching Award, and the Professor's Teaching Award for the Preclinical Sciences. He is coauthor, with Stephen J. Lippard, of the textbook *Principles of Bioinorganic Chemistry*.

JOHN L. TYMOCZKO is Towsley Professor of Biology at Carleton College, where he has taught since 1976. He currently teaches Biochemistry, Biochemistry Laboratory, Oncogenes and the Molecular Biology of Cancer, and Exercise Biochemistry and coteaches an introductory course, Energy Flow in Biological Systems. Professor Tymoczko received his B.A. from the University of Chicago in 1970 and his Ph.D. in Biochemistry from the University of Chicago with Shutsung Liao at the Ben May Institute for Cancer Research. He then had a postdoctoral position with Hewson Swift of the Department of Biology at the University of Chicago. The focus of his research has been on steroid receptors, ribonucleoprotein particles, and proteolytic processing enzymes.

GREGORY J. GATTO, JR., received his A.B. degree in Chemistry from Princeton University, where he worked with Martin F. Semmelhack and was awarded the Everett S. Wallis Prize in Organic Chemistry. In 2003, he received his M.D. and Ph.D. degrees from the Johns Hopkins University School of Medicine, where he studied the structural biology of peroxisomal targeting signal recognition with Jeremy M. Berg and received the Michael A. Shanoff Young Investigator Research Award. He completed a postdoctoral fellowship in 2006 with Christopher T. Walsh at Harvard Medical School, where he studied the biosynthesis of the macrolide immunosuppressants. He is currently a Senior Scientific Investigator in the Heart Failure Discovery Performance Unit at GlaxoSmithKline.

LUBERT STRYER is Winzer Professor of Cell Biology, Emeritus, in the School of Medicine and Professor of Neurobiology, Emeritus, at Stanford University, where he has been on the faculty since 1976. He received his M.D. from Harvard Medical School. Professor Stryer has received many awards for his research on the interplay of light and life, including the Eli Lilly Award for Fundamental Research in Biological Chemistry, the Distinguished Inventors Award of the Intellectual Property Owners' Association, and election to the National Academy of Sciences and the American Philosophical Society. He was awarded the National Medal of Science in 2006. The publication of his first edition of *Biochemistry* in 1975 transformed the teaching of biochemistry.

PREFACE

For several generations of students and teachers, *Biochemistry* has been an invaluable resource, presenting the concepts and details of molecular structure, metabolism, and laboratory techniques in a streamlined and engaging way. *Biochemistry*'s success in helping students learn the subject for the first time is built on a number of hallmark features:

- **Clear writing and simple illustrations**. The language of biochemistry is made as accessible as possible for students learning the subject for the first time. To complement the straightforward language and organization of concepts in the text, figures illustrate a single concept at a time to help students see main points without the distraction of excess detail.
- **Physiological relevance**. It has always been our goal to help students connect biochemistry to their own lives on a variety of scales. Pathways and processes are presented in a physiological context so students can see how biochemistry works in the body and under different conditions, and Clinical Application sections in every chapter show students how the concepts they are studying impact human health. The eighth edition includes a number of new Clinical Application sections based on recent discoveries in biochemistry and health. (For a full list, see p. xi)
- **Evolutionary perspective**. Discussions of evolution are woven into the narrative of the text, just as evolution shapes every pathway and molecular structure described in the text. Molecular Evolution sections highlight important milestones in the evolution of life as a way to provide context for the processes and molecules being discussed. (For a full list, see p. x)
- **Problem-solving practice**. Every chapter of *Biochemistry* provides numerous opportunities for students to practice problem-solving skills and apply the concepts described in the text. End-of-chapter problems are divided into three categories to address different problem-solving skills: Mechanism problems ask students to suggest or describe a chemical mechanism; Data interpretation problems ask students to draw conclusions from data taken from real research papers; and chapter integration problems require students to connect concepts from across chapters. Further problem-solving practice is provided online, on the *Biochemistry* LaunchPad. (For more details on LaunchPad resources, see p. viii)
- **A variety of molecular structures**. All molecular structures in the book, with few exceptions, have been selected and rendered by Jeremy Berg and Gregory Gatto to emphasize the aspect of structure most important to the topic at hand. Students are introduced to realistic renderings of molecules through a molecular model "primer" in the appendices to Chapters 1 and 2 so they are well-equipped to recognize and interpret the structures throughout the book. Figure legends direct students explicitly to the key features of a model, and often include PDB numbers so the reader can access the file used in generating the structure from the Protein Data Bank website (www.pdb.org). Students

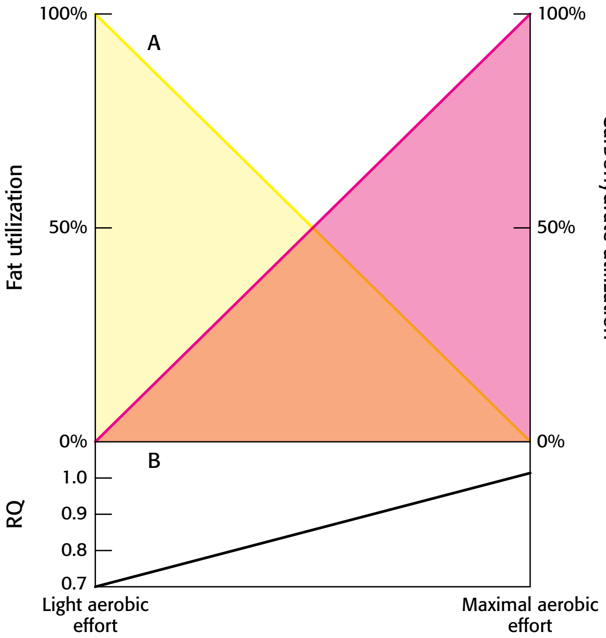


Figure 27.12 An idealized representation of fuels use as a function of aerobic exercise intensity. (A) With increased exercise intensity, the use of fats as fuels falls as the utilization of glucose increases. (B) The respiratory quotient (RQ) measures the alteration in fuel use.

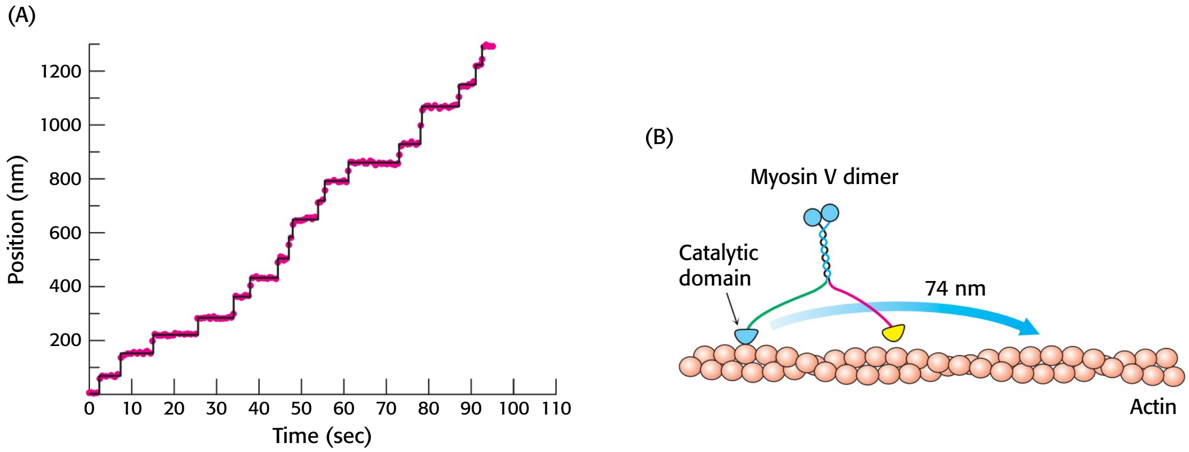


Figure 9.48 Single molecule motion. (A) A trace of the position of a single dimeric myosin V molecule as it moves across a surface coated with actin filaments. (B) A model of how the dimeric molecule moves in discrete steps with an average size of 74 ± 5 nm. [Data from A. Yildiz et al., Science 300(5628)2061–2065, 2003.]

can explore molecular structures further online through the Living Figures, in which they can rotate 3D models of molecules and view alternative renderings.

In this revision of *Biochemistry*, we focused on building on the strengths of the previous editions to present biochemistry in an even more clear and streamlined manner, as well as incorporating exciting new advances from the field. Throughout the book, we have updated explanations of basic concepts and bolstered them with examples from new research. Some new topics that we present in the eighth edition include:

- Environmental factors that influence human biochemistry (Chapter 1)
- Genome editing (Chapter 5)
- Horizontal gene transfer events that may explain unexpected branches of the evolutionary tree (Chapter 6)
- Penicillin irreversibly inactivating a key enzyme in bacterial cell-wall synthesis (Chapter 8)
- Scientists watching single molecules of myosin move (Chapter 9)
- Glycosylation functions in nutrient sensing (Chapter 11)
- The structure of a SNARE complex (Chapter 12)
- The mechanism of ABC transporters (Chapter 13)
- The structure of the gap junction (Chapter 13)
- The structural basis for activation of the β-adrenergic receptor (Chapter 14)
- Excessive fructose consumption can lead to pathological conditions (Chapter 16)
- Alterations in the glycolytic pathway by cancer cells (Chapter 16)
- Regulation of mitochondrial ATP synthase (Chapter 18)
- Control of chloroplast ATP synthase (Chapter 19)
- Activation of rubisco by rubisco activase (Chapter 20)

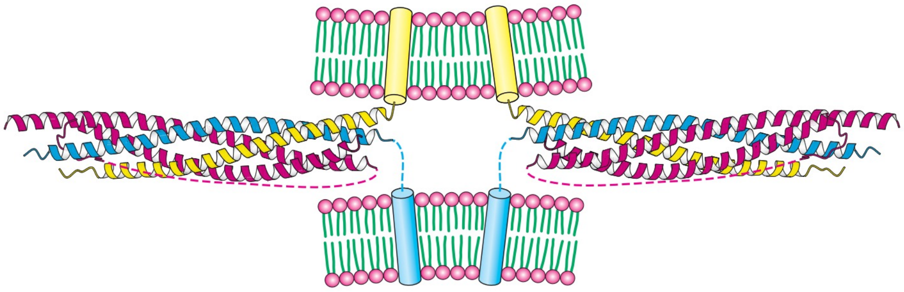


Figure 12.39 SNARE complexes initiate membrane fusion. The SNARE protein synaptobrevin (yellow) from one membrane forms a tight four-helical bundle with the corresponding SNARE proteins syntaxin-1 (blue) and SNAP25 (red) from a second membrane. The complex brings the membranes close together, initiating the fusion event. [Drawn from 1SFC.pdb.]

- The role of the pentose phosphate pathway in rapid cell growth (Chapter 20)
- Biochemical characteristics of muscle fiber types (Chapter 21)
- Alteration of fatty acid metabolism in tumor cells (Chapter 22)
- Biochemical basis of neurological symptoms of phenylketonuria (Chapter 24)
- Ribonucleotide reductase as a chemotherapeutic target (Chapter 25)
- The role of excess choline in the development of heart disease (Chapter 26)
- Cycling of the LDL receptor is regulated (Chapter 26)
- The role of ceramide metabolism in stimulating tumor growth (Chapter 26)
- The extraordinary power of DNA repair systems illustrated by *Deinococcus radiodurans* (Chapter 28)
- The structural details of ligand binding by TLRs (Chapter 34)

LaunchPad

All of the new media resources for *Biochemistry* will be available in our new system.

www.macmillanhighered.com/launchpad/berg8e

LaunchPad is a dynamic, fully integrated learning environment that brings together all of our teaching and learning resources in one place. It also contains the fully interactive **e-Book** and other newly updated resources for students and instructors, including the following:

- **NEW Case Studies** are a series of biochemistry case studies you can integrate into your course. Each case study gives students practice in working with data, developing critical thinking skills, connecting topics, and applying knowledge to real scenarios. We also provide instructional guidance with each case study (with suggestions on how to use the case in the classroom) and aligned assessment questions for quizzes and exams.
- **Newly Updated Clicker Questions** allow instructors to integrate active learning in the classroom and to assess students' understanding of key concepts during lectures. Available in Microsoft Word and PowerPoint (PPT).
- **Newly Updated Lecture PowerPoints** have been developed to minimize preparation time for new users of the book. These files offer suggested lectures including key illustrations and summaries that instructors can adapt to their teaching styles.
  - **Updated Layered PPTs** deconstruct key concepts, sequences, and processes from the textbook images, allowing instructors to present complex ideas step-by-step.
  - **Updated Textbook Images and Tables** are offered as high-resolution JPEG files. Each image has been fully optimized to increase type sizes and adjust color saturation. These images have been tested in a large lecture hall to ensure maximum clarity and visibility.
  - **The Clinical Companion**, by Gregory Raner, The University of North Carolina at Greensboro and Douglas Root, University of North Texas, applies concepts that students have learned in the book to novel medical situations. Students read clinical case studies and use basic biochemistry concepts to solve the medical mysteries, applying and reinforcing what they learn in lecture and from the book.
  - **Hundreds of self-graded practice problems** allow students to test their understanding of concepts explained in the text, with immediate feedback.
  - **The Metabolic Map** helps students understand the principles and applications of the core metabolic pathways. Students can work through guided tutorials with embedded assessment questions, or explore the Metabolic Map on their own using the dragging and zooming functionality of the map.
  - **Jmol tutorials** by Jeffrey Cohlberg, California State University at Long Beach, teach students

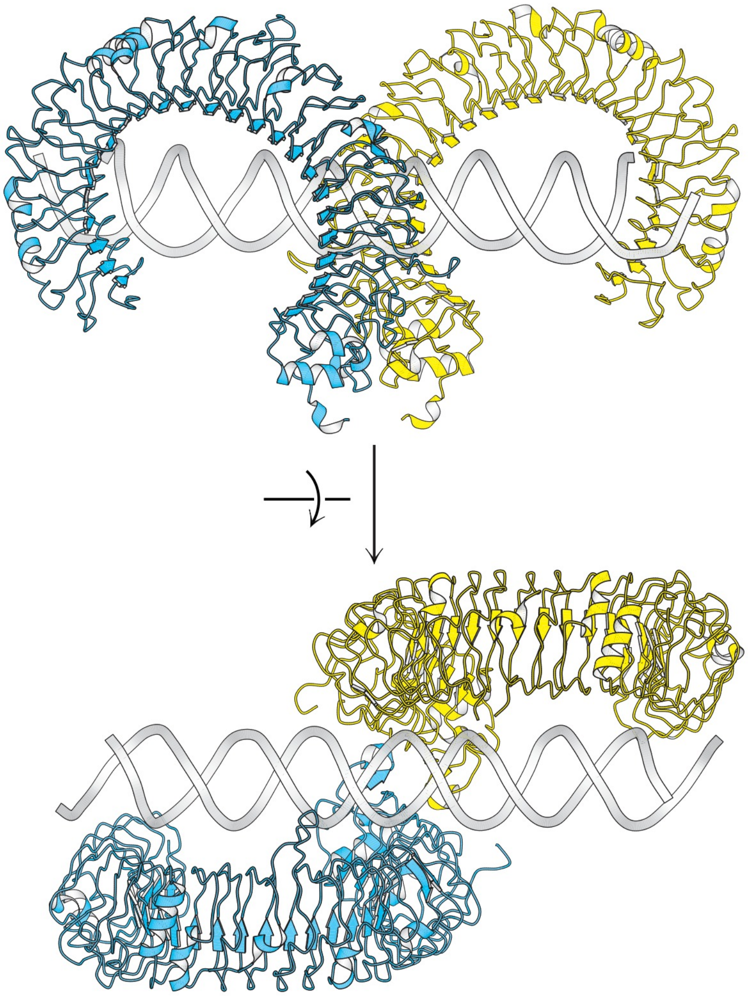


Figure 34.3 Recognition of a PAMP by a Toll-like receptor. The structure of TLR3 bound to its PAMP, a fragment of double-stranded RNA, as seen from the side (top) and from above (bottom). *Notice* that the PAMP induces receptor dimerization by binding the surfaces on the side of each of the extracellular domains. [Drawn from 3CIY.pdb].

how to create models of proteins in Jmol based on data from the Protein Data Bank. By working through the tutorial and answering assessment questions at the end of each exercise, students learn to use this important database and fully realize the relationships between the structure and function of enzymes.

- **Living figures** allow students to explore protein structure in 3-D. Students can zoom and rotate the "live" structures to get a better understanding of their three-dimensional nature and can experiment with different display styles (space-filling, ball-and-stick, ribbon, backbone) by means of a user-friendly interface.
- **Concept-based tutorials** by Neil D. Clarke help students build an intuitive understanding of some of the more difficult concepts covered in the textbook.
- **Animated techniques** help students grasp experimental techniques used for exploring genes and proteins.
- **NEW animations** show students biochemical processes in motion. The eighth edition includes many new animations.
- **Online end-of-chapter questions** are assignable and self-graded multiple-choice versions of the end-of-chapter questions in the book, giving students a way to practice applying chapter content in an online environment.
- **Flashcards** are an interactive tool that allows students to study key terms from the book.
- **LearningCurve** is a self-assessment tool that helps students evaluate their progress. Students can test their understanding by taking an online multiple-choice quiz provided for each chapter, as well as a general chemistry review.

Updated Student Companion

[1-4641-8803-3]

For each chapter of the textbook, the *Student Companion* includes:

- **Chapter Learning Objectives and Summary**
- **Self-Assessment Problems,** including multiple-choice, short-answer, matching questions, and challenge problems, and their answers
- **Expanded Solutions** to end-of-chapter problems in the textbook

MOLECULAR EVOLUTION

This icon signals the start of the many discussions that highlight protein commonalities or other molecular evolutionary insights.

Only L amino acids make up proteins (p. 29)
Why this set of 20 amino acids? (p. 35)
Sickle-cell trait and malaria (p. 206)
Additional human globin genes (p. 208)
Catalytic triads in hydrolytic enzymes (p. 258)
Major classes of peptide-cleaving enzymes (p. 260)
Common catalytic core in type II restriction enzymes (p. 275)
P-loop NTPase domains (p. 280)
Conserved catalytic core in protein kinases (p. 298)
Why do different human blood types exist? (p. 331)
Archaeal membranes (p. 346)
Ion pumps (p. 370)
P-type ATPases (p. 374)
ATP-binding cassettes (p. 374)
Sequence comparisons of Na^+ and Ca^{2+} channels (p. 382)
Small G proteins (p. 414)
Metabolism in the RNA world (p. 444)
Why is glucose a prominent fuel? (p. 451)
NAD^+ binding sites in dehydrogenases (p. 465)
Isozymic forms of lactate dehydrogenase (p. 487)
Evolution of glycolysis and gluconeogenesis (p. 487)
The α-ketoglutarate dehydrogenase complex (p. 505)
Domains of succinyl CoA synthetase (p. 507)
Evolution of the citric acid cycle (p. 516)
Mitochondrial evolution (p. 525)
Conserved structure of cytochrome *c* (p. 541)
Common features of ATP synthase and G proteins (p. 548)
Pigs lack uncoupling protein 1 (UCP-1) and brown fat (p. 556)
Related uncoupling proteins (p. 556)
Chloroplast evolution (p. 568)
Evolutionary origins of photosynthesis (p. 584)
Evolution of the C4 pathway (p. 601)
The relationship of the Calvin cycle and the pentose phosphate pathway (p. 610)
Increasing sophistication of glycogen phosphorylase regulation (p. 629)
Glycogen synthase is homologous to glycogen phosphorylase (p. 631)
A recurring motif in the activation of carboxyl groups (p. 649)
Prokaryotic counterparts of the ubiquitin pathway and the proteasome (p. 686)
A family of pyridoxal-dependent enzymes (p. 692)
Evolution of the urea cycle (p. 696)
The P-loop NTPase domain in nitrogenase (p. 716)
Conserved amino acids in transaminases determine amino acid chirality (p. 721)
Feedback inhibition (p. 731)
Recurring steps in purine ring synthesis (p. 749)
Ribonucleotide reductases (p. 755)
Increase in urate levels during primate evolution (p. 761)
Deinococcus radiodurans illustrates the power of DNA repair systems (p. 828)
DNA polymerases (p. 829)
Thymine and the fidelity of the genetic message (p. 849)
Sigma factors in bacterial transcription (p. 865)
Similarities in transcription between archaea and eukaryotes (p. 876)
Evolution of spliceosome-catalyzed splicing (p. 888)
Classes of aminoacyl-tRNA synthetases (p. 901)
Composition of the primordial ribosome (p. 903)
Homologous G proteins (p. 908)
A family of proteins with common ligand-binding domains (p. 930)
The independent evolution of DNA-binding sites of regulatory proteins (p. 931)
Key principles of gene regulation are similar in bacteria and archaea (p. 937)
CpG islands (p. 949)
Iron-response elements (p. 955)
miRNAs in gene evolution (p. 957)
The odorant-receptor family (p. 963)
Photoreceptor evolution (p. 973)
The immunoglobulin fold (p. 988)
Relationship of tubulin to prokaryotic proteins (p. 1023)

CLINICAL APPLICATIONS

This icon signals the start of a clinical application in the text. Additional, briefer clinical correlations appear in the text as appropriate.

Osteogenesis imperfecta (p. 46)
Protein-misfolding diseases (p. 56)
Protein modification and scurvy (p. 57)
Antigen/antibody detection with ELISA (p. 82)
Synthetic peptides as drugs (p. 92)
PCR in diagnostics and forensics (p.142)
Gene therapy (p. 164)
Aptamers in biotechnology and medicine (p. 187)
Functional magnetic resonance imaging (p. 193)
2,3-BPG and fetal hemoglobin (p. 201)
Carbon monoxide poisoning (p. 201)
Sickle-cell anemia (p. 205)
Thalassemia (p. 207)
Aldehyde dehydrogenase deficiency (p. 228)
Action of penicillin (p. 239)
Protease inhibitors (p. 263)
Carbonic anhydrase and osteopetrosis (p. 264)
Isozymes as a sign of tissue damage (p. 293)
Trypsin inhibitor helps prevent pancreatic damage (p. 302)
Emphysema (p. 303)
Blood clotting involves a cascade of zymogen activations (p. 303)
Vitamin K (p. 306)
Antithrombin and hemorrhage (p. 307)
Hemophilia (p.308)
Monitoring changes in glycosylated hemoglobin (p. 321)
Erythropoietin (p. 327)
Hurler disease (p. 327)
Mucins (p. 329)
Blood groups (p. 331)
I-cell disease (p. 332)
Influenza virus binding (p. 335)
Clinical applications of liposomes (p. 349)
Aspirin and ibuprofen (p. 353)
Digitalis and congestive heart failure (p. 373)
Multidrug resistance (p. 374)
Long QT syndrome (p. 388)
Signal-transduction pathways and cancer (p. 416)
Monoclonal antibodies as anticancer drugs (p. 416)
Protein kinase inhibitors as anticancer drugs (p. 417)
G-proteins, cholera and whooping cough (p. 417)
Vitamins (p. 438)
Triose phosphate isomerase deficiency (p. 454)
Excessive fructose consumption (p. 466)
Lactose intolerance (p. 467)
Galactosemia (p. 468)
Aerobic glycolysis and cancer (p. 474)
Phosphatase deficiency (p. 512)
Defects in the citric acid cycle and the development of cancer (p. 513)
Beriberi and mercury poisoning (p. 515)
Frataxin mutations cause Friedreich's ataxia (p. 531)
Reactive oxygen species (ROS) are implicated in a variety of diseases (p. 539)
ROS may be important in signal transduction (p. 540)
IF1 overexpression and cancer (p. 554)
Brown adipose tissue (p. 555)
Mild uncouplers sought as drugs (p.557)
Mitochondrial diseases (p. 557)
Glucose 6-phosphate dehydrogenase deficiency causes drug-induced hemolytic anemia (p. 610)
Glucose 6-phosphate dehydrogenase deficiency protects against malaria (p. 612)
Developing drugs for type 2 diabetes (p. 636)
Glycogen-storage diseases (p. 637)
Chanarin-Dorfman syndrome (p. 648)
Carnitine deficiency (p. 650)
Zellweger syndrome (p. 657)
Diabetic ketosis (p. 659)
Ketogenic diets to treat epilepsy (p. 660)
Some fatty acids may contribute to pathological conditions (p. 661)
The use of fatty acid synthase inhibitors as drugs (p. 667)
Effects of aspirin on signaling pathways (p. 669)
Diseases resulting from defects in transporters of amino acids (p. 682)
Diseases resulting from defects in E3 proteins (p. 685)
Drugs target the ubiquitin-proteasome system (p.687)
Using proteasome inhibitors to treat tuberculosis (p. 687)
Blood levels of aminotransferases indicate liver damage (p. 691)
Inherited defects of the urea cycle (hyperammonemia) (p. 697)
Alcaptonuria, maple syrup urine disease, and phenylketonuria (p. 705)

High homocysteine levels and vascular disease (p. 726)
Inherited disorders of porphyrin metabolism (p. 737)
Anticancer drugs that block the synthesis of thymidylate (p. 757)
Ribonucleotide reductase is a target for cancer therapy (p. 759)
Adenosine deaminase and severe combined immunodeficiency (p. 760)
Gout (p. 761)
Lesch–Nyhan syndrome (p. 761)
Folic acid and spina bifida (p. 762)
Enzyme activation in some cancers to generate phosphocholine (p. 770)
Excess choline and heart disease (p. 771)
Gangliosides and cholera (p. 773)
Second messengers derived from sphingolipids and diabetes (p. 773)
Respiratory distress syndrome and Tay–Sachs disease (p. 774)
Ceramide metabolism stimulates tumor growth (p. 775)
Phosphatidic acid phosphatase and lipodystrophy (p. 776)
Hypercholesterolemia and atherosclerosis (p. 784)
Mutations in the LDL receptor (p. 785)
LDL receptor cycling is regulated (p. 787)
The role of HDL in protecting against arteriosclerosis (p. 787)
Clinical management of cholesterol levels (p. 788)
Bile salts are derivatives of cholesterol (p. 789)
The cytochrome P450 system is protective (p. 791)
A new protease inhibitor also inhibits a cytochrome P450 enzyme (p. 792)
Aromatase inhibitors in the treatment of breast and ovarian cancer (p. 794)
Rickets and vitamin D (p. 795)
Caloric homeostasis is a means of regulating body weight (p. 802)
The brain plays a key role in caloric homeostasis (p. 804)
Diabetes is a common metabolic disease often resulting from obesity (p. 807)
Exercise beneficially alters the biochemistry of cells (p. 813)
Food intake and starvation induce metabolic changes (p. 816)
Ethanol alters energy metabolism in the liver (p. 819)
Antibiotics that target DNA gyrase (p. 839)
Blocking telomerase to treat cancer (p. 845)
Huntington disease (p. 850)
Defective repair of DNA and cancer (p. 850)
Detection of carcinogens (Ames test) (p. 852)
Translocations can result in diseases (p. 855)
Antibiotic inhibitors of transcription (p. 869)
Burkitt lymphoma and B-cell leukemia (p. 876)
Diseases of defective RNA splicing (p. 884)
Vanishing white matter disease (p. 913)
Antibiotics that inhibit protein synthesis (p. 914)
Diphtheria (p. 914)
Ricin, a lethal protein-synthesis inhibitor (p. 915)
Induced pluripotent stem cells (p. 947)
Anabolic steroids (p. 951)
Color blindness (p. 974)
The use of capsaicin in pain management (p. 978)
Immune-system suppressants (p. 994)
MHC and transplantation rejection (p. 1002)
AIDS (p. 1003)
Autoimmune diseases (p. 1005)
Immune system and cancer (p. 1005)
Vaccines (p. 1006)
Charcot-Marie-Tooth disease (p. 1022)
Taxol (p. 1023)

ACKNOWLEDGMENTS

Writing a popular textbook is both a challenge and an honor. Our goal is to convey to our students our enthusiasm and understanding of a discipline to which we are devoted. They are our inspiration. Consequently, not a word was written or an illustration constructed without the knowledge that bright, engaged students would immediately detect vagueness and ambiguity. We also thank our colleagues who supported, advised, instructed, and simply bore with us during this arduous task. We are grateful to our colleagues throughout the world who patiently answered our questions and shared their insights into recent developments.

We also especially thank those who served as reviewers for this new edition. Their thoughtful comments, suggestions, and encouragement have been of immense help to us in maintaining the excellence of the preceding editions. These reviewers are:

Paul Adams
University of Arkansas, Fayetteville
Kevin Ahern
Oregon State University
Zulfiqar Ahmad
A.T. Still University of Health Sciences
Young-Hoon An
Wayne State University
Richard Amasino
University of Wisconsin
Kenneth Balazovich
University of Michigan
Donald Beitz
Iowa State University
Matthew Berezuk
Azusa Pacific University
Melanie Berkmen
Suffolk University
Steven Berry
University of Minnesota, Duluth
Loren Bertocci
Marian University
Mrinal Bhattacharjee
Long Island University
Elizabeth Blinstrup-Good
University of Illinois
Brian Bothner
Montana State University
Mark Braiman
Syracuse University
David Brown
Florida Gulf Coast University
Donald Burden
Middle Tennessee State University
Nicholas Burgis
Eastern Washington University
W. Malcom Byrnes
Howard University College of Medicine
Graham Carpenter
Vanderbilt University School of Medicine
John Cogan
The Ohio State University
Jeffrey Cohlberg
California State University, Long Beach
David Daleke
Indiana University
John DeBanzie
Northeastern State University
Cassidy Dobson
St. Cloud State University
Donald Doyle
Georgia Institute of Technology
Ludeman Eng
Virginia Tech
Caryn Evilia
Idaho State University
Kirsten Fertuck
Northeastern University
Brent Feske
Armstrong Atlantic University
Patricia Flatt
Western Oregon University
Wilson Francisco
Arizona State University
Gerald Frenkel
Rutgers University
Ronald Gary
University of Nevada, Las Vegas
Eric R. Gauthier
Laurentian University
Glenda Gillaspy
Virginia Tech
James Gober
UCLA
Christina Goode
California State University, Fullerton
Nina Goodey
Montclair State University
Eugene Grgory
Virginia Tech
Robert Grier
Atlanta Metropolitan State College
Neena Grover
Colorado College
Paul Hager
East Carolina University
Ann Hagerman
Miami University
Mary Hatcher-Skeers
Scripps College
Diane Hawley
University of Oregon
Blake Hill
Medical College of Wisconsin
Pui Ho
Colorado State University
Charles Hoogstraten
Michigan State University
Frans Huijing
University of Miami
Kathryn Huisinga
Malone University
Cristi Junnes
Rocky Mountain College
Lori Isom
University of Central Arkansas
Nitin Jain
University of Tennessee
Blythe Janowiak
Saint Louis University
Gerwald Jogl
Brown University
Kelly Johanson
Xavier University of Louisiana
Jerry Johnson
University of Houston-Downtown
Todd Johnson
Weber State University
David Josephy
University of Guelph
Michael Kalafatis
Cleveland State University
Marina Kazakevich
University of Massachusetts-Dartmouth
Jong Kim
Alabama A&M University

Sung-Kun Kim
Baylor University

Roger Koeppe
University of Arkansas, Fayetteville

Dmitry Kolpashchikov
University of Central Florida

Min-Hao Kuo
Michigan State University

Isabel Larraza
North Park University

Mark Larson
Augustana College

Charles Lawrence
Montana State University

Pan Li
State University of New York, Albany

Darlene Loprete
Rhodes College

Greg Marks
Carroll University

Michael Massiah
George Washington University

Keri McFarlane
Northern Kentucky University

Michael Mendenhall
University of Kentucky

Stephen Mills
University of San Diego

Smita Mohanty
Auburn University

Debra Moriarity
University of Alabama, Huntsville

Stephen Munroe
Marquette University

Jeffrey Newman
Lycoming College

William Newton
Virginia Tech

Alfred Nichols
Jacksonville State University

Brian Nichols
University of Illinois, Chicago

Allen Nicholson
Temple University

Brad Nolen
University of Oregon

Pamela Osenkowski
Loyola University, Chicago

Xiaping Pan
East Carolina University

Stefan Paula
Northern Kentucky University

David Pendergrass
University of Kansas-Edwards

Wendy Pogozelski
State University of New York, Geneseo

Gary Powell
Clemson University

Geraldine Prody
Western Washington University

Joseph Provost
University of San Diego

Greg Raner
University of North Carolina, Greensboro

Tanea Reed
Eastern Kentucky University

Christopher Reid
Bryant University

Denis Revie
California Lutheran University

Douglas Root
University of North Texas

Johannes Rudolph
University of Colorado

Brian Sato
University of California, Irvine

Glen Sauer
Fairfield University

Joel Schildbach
Johns Hopkins University

Stylianos Scordilis
Smith College

Ashikh Seethy
Maulana Azad Medical College, New Delhi

Lisa Shamansky
California State University, San Bernardino

Bethel Sharma
Sewanee: University of the South

Nicholas Silvaggi
University of Wisconsin-Milwaukee

Kerry Smith
Clemson University

Narashima Sreerama
Colorado State University

Wesley Stites
University of Arkansas

Jon Stoltzfus
Michigan State University

Gerald Stubbs
Vanderbilt University

Takita Sumter
Winthrop University

Anna Tan-Wilson
State University of New York, Binghamton

Steven Theg
University of California, Davis

Marc Tischler
University of Arizona

Ken Traxler
Bemidji State University

Brian Trewyn
Colorado School of Mines

Vishwa Trivedi
Bethune Cookman University

Panayiotis Vacratsis
University of Windsor

Peter van der Geer
San Diego State University

Jeffrey Voigt
Albany College of Pharmacy and Health Sciences

Grover Waldrop
Louisiana State University

Xuemin Wang
University of Missouri

Yuqi Wang
Saint Louis University

Rodney Weilbaecher
Southern Illinois University

Kevin Williams
Western Kentucky University

Laura Zapanta
University of Pittsburgh

Brent Znosko
Saint Louis University

We have been working with the people at W. H. Freeman/Macmillan Higher Education for many years now, and our experiences have always been enjoyable and rewarding. Writing and producing the eighth edition of *Biochemistry* confirmed our belief that they are a wonderful publishing team and we are honored to work with them. Our Macmillan colleagues have a knack for undertaking stressful, but exhilarating, projects and reducing the stress without reducing the exhilaration and a remarkable ability to coax without ever nagging. We have many people to thank for this experience, some of whom are first timers to the *Biochemistry* project.

We are delighted to work with Senior Acquisitions Editor, Lauren Schultz, for the first time. She was unfailing in her enthusiasm and generous with her support. Another new member of the team was our developmental editor, Irene Pech. We have had the pleasure of working with a number of outstanding developmental editors over the years, and Irene continues this tradition. Irene is thoughtful, insightful, and very efficient at identifying aspects of our writing and figures that were less than clear. Lisa Samols, a former developmental editor, served as a consultant, archivist for previous editions, and a general source of publishing knowledge. Senior Project Editor Deni Showers, with Sherrill Redd, managed the flow of the entire project, from copyediting through bound book, with admirable efficiency. Irene Vartanoff and Mercy Heston, our manuscript editors, enhanced the literary consistency and clarity of the text. Vicki Tomaselli, Design Manager, produced a design and layout that makes the book uniquely attractive while still emphasizing its ties to past editions. Photo Editor Christine Buese and Photo Researcher Jacalyn Wong found the photographs that we hope make the text not only more inviting, but also fun to look through. Janice Donnola, Illustration Coordinator, deftly directed the rendering of new illustrations. Paul Rohloff, Production Coordinator, made sure that the significant difficulties of scheduling, composition, and manufacturing were smoothly overcome. Amanda Dunning and Donna Brodman did a wonderful job in their management of the media program. In addition, Amanda ably coordinated the print supplements plan. Special thanks also to editorial assistants Shannon Moloney and Nandini Ahuja. Sandy Lindelof, Executive Marketing Manager, enthusiastically introduced this newest edition of *Biochemistry* to the academic world. We are deeply appreciative of Craig Bleyer and his sales staff for their support. Without their able and enthusiastic presentation of our text to the academic community, all of our efforts would be in vain. We also wish to thank Kate Ahr Parker, Publisher, for her encouragement and belief in us.

Thanks also to our many colleagues at our own institutions as well as throughout the country who patiently answered our questions and encouraged us on our quest. Finally, we owe a debt of gratitude to our families—our wives, Wendie Berg, Alison Unger, and Megan Williams, and our children, especially Timothy and Mark Gatto. Without their support, comfort, and understanding, this endeavor could never have been undertaken, let alone successfully completed.

BRIEF CONTENTS

Part I THE MOLECULAR DESIGN OF LIFE

- 1 Biochemistry: An Evolving Science 1
- 2 Protein Composition and Structure 27
- 3 Exploring Proteins and Proteomes 65
- 4 DNA, RNA, and the Flow of Genetic Information 105
- 5 Exploring Genes and Genomes 135
- 6 Exploring Evolution and Bioinformatics 169
- 7 Hemoglobin: Portrait of a Protein in Action 191
- 8 Enzymes: Basic Concepts and Kinetics 215
- 9 Catalytic Strategies 251
- 10 Regulatory Strategies 285
- 11 Carbohydrates 315
- 12 Lipids and Cell Membranes 341
- 13 Membrane Channels and Pumps 367
- 14 Signal-Transduction Pathways 397

Part II TRANSDUCING AND STORING ENERGY

- 15 Metabolism: Basic Concepts and Design 423
- 16 Glycolysis and Gluconeogenesis 449
- 17 The Citric Acid Cycle 495
- 18 Oxidative Phosphorylation 523
- 19 The Light Reactions of Photosynthesis 565
- 20 The Calvin Cycle and the Pentose Phosphate Pathway 589
- 21 Glycogen Metabolism 617
- 22 Fatty Acid Metabolism 643
- 23 Protein Turnover and Amino Acid Catabolism 681

Part III SYNTHESIZING THE MOLECULES OF LIFE

- 24 The Biosynthesis of Amino Acids 713
- 25 Nucleotide Biosynthesis 743
- 26 The Biosynthesis of Membrane Lipids and Steroids 767
- 27 The Integration of Metabolism 801
- 28 DNA Replication, Repair, and Recombination 827
- 29 RNA Synthesis and Processing 859
- 30 Protein Synthesis 893
- 31 The Control of Gene Expression in Prokaryotes 925
- 32 The Control of Gene Expression in Eukaryotes 941

Part IV RESPONDING TO ENVIRONMENTAL CHANGES

- 33 Sensory Systems 961
- 34 The Immune System 981
- 35 Molecular Motors 1011
- 36 Drug Development 1033

CONTENTS

Preface v

Part I THE MOLECULAR DESIGN OF LIFE 1

CHAPTER 1 Biochemistry: An Evolving Science 1

- **1.1** Biochemical Unity Underlies Biological Diversity 1
- **1.2** DNA Illustrates the Interplay Between Form and Function 4
  - DNA is constructed from four building blocks 4
  - Two single strands of DNA combine to form a double helix 5
  - DNA structure explains heredity and the storage of information 5
- **1.3** Concepts from Chemistry Explain the Properties of Biological Molecules 6
  - The formation of the DNA double helix as a key example 6
  - The double helix can form from its component strands 6
  - Covalent and noncovalent bonds are important for the structure and stability of biological molecules 6
  - The double helix is an expression of the rules of chemistry 9
  - The laws of thermodynamics govern the behavior of biochemical systems 10
  - Heat is released in the formation of the double helix 12
  - Acid–base reactions are central in many biochemical processes 13
  - Acid–base reactions can disrupt the double helix 14
  - Buffers regulate pH in organisms and in the laboratory 15
- **1.4** The Genomic Revolution Is Transforming Biochemistry, Medicine, and Other Fields 17
  - Genome sequencing has transformed biochemistry and other fields 17
  - Environmental factors influence human biochemistry 20
  - Genome sequences encode proteins and patterns of expression 21

APPENDIX: Visualizing Molecular Structures I: Small Molecules 22

CHAPTER 2 Protein Composition and Structure 27

- **2.1** Proteins Are Built from a Repertoire of 20 Amino Acids 29
- **2.2** Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains 35
  - Proteins have unique amino acid sequences specified by genes 37
  - Polypeptide chains are flexible yet conformationally restricted 38
- **2.3** Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops 40
  - The alpha helix is a coiled structure stabilized by intrachain hydrogen bonds 40
  - Beta sheets are stabilized by hydrogen bonding between polypeptide strands 42

Polypeptide chains can change direction by making reverse turns and loops 44
Fibrous proteins provide structural support for cells and tissues 44
2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores 46
2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures 48
2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure 49
Amino acids have different propensities for forming α helices, β sheets, and turns 51
Protein folding is a highly cooperative process 52
Proteins fold by progressive stabilization of intermediates rather than by random search 53
Prediction of three-dimensional structure from sequence remains a great challenge 54
Some proteins are inherently unstructured and can exist in multiple conformations 55
Protein misfolding and aggregation are associated with some neurological diseases 56
Protein modification and cleavage confer new capabilities 57
APPENDIX: Visualizing Molecular Structures II: Proteins 61

CHAPTER 3 Exploring Proteins and Proteomes 65

The proteome is the functional representation of the genome 66
3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function 66
The assay: How do we recognize the protein that we are looking for? 67
Proteins must be released from the cell to be purified 67
Proteins can be purified according to solubility, size, charge, and binding affinity 68
Proteins can be separated by gel electrophoresis and displayed 71
A protein purification scheme can be quantitatively evaluated 75
Ultracentrifugation is valuable for separating biomolecules and determining their masses 76
Protein purification can be made easier with the use of recombinant DNA technology 78
3.2 Immunology Provides Important Techniques with Which to Investigate Proteins 79
Antibodies to specific proteins can be generated 79
Monoclonal antibodies with virtually any desired specificity can be readily prepared 80
Proteins can be detected and quantified by using an enzyme-linked immunosorbent assay 82
Western blotting permits the detection of proteins separated by gel electrophoresis 83
Fluorescent markers make the visualization of proteins in the cell possible 84
3.3 Mass Spectrometry Is a Powerful Technique for the Identification of Peptides and Proteins 85
Peptides can be sequenced by mass spectrometry 87
Proteins can be specifically cleaved into small peptides to facilitate analysis 88
Genomic and proteomic methods are complementary 89
The amino acid sequence of a protein provides valuable information 90
Individual proteins can be identified by mass spectrometry 91
3.4 Peptides Can Be Synthesized by Automated Solid-Phase Methods 92
3.5 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy 95
X-ray crystallography reveals three-dimensional structure in atomic detail 95
Nuclear magnetic resonance spectroscopy can reveal the structures of proteins in solution 97

CHAPTER 4 DNA, RNA, and the Flow of Genetic Information 105

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone 106
RNA and DNA differ in the sugar component and one of the bases 106
Nucleotides are the monomeric units of nucleic acids 107
DNA molecules are very long and have directionality 108
4.2 A Pair of Nucleic Acid Strands with Complementary Sequences Can Form a Double-Helical Structure 109
The double helix is stabilized by hydrogen bonds and van der Waals interactions 109
DNA can assume a variety of structural forms 111
Z-DNA is a left-handed double helix in which backbone phosphates zigzag 112
Some DNA molecules are circular and supercoiled 113
Single-stranded nucleic acids can adopt elaborate structures 113
4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information 114
Differences in DNA density established the validity of the semiconservative replication hypothesis 115
The double helix can be reversibly melted 116
4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates 117
DNA polymerase catalyzes phosphodiester-bridge formation 117
The genes of some viruses are made of RNA 118
4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules 119
Several kinds of RNA play key roles in gene expression 119

All cellular RNA is synthesized by RNA polymerases 120
RNA polymerases take instructions from DNA templates 121
Transcription begins near promoter sites and ends at terminator sites 122
Transfer RNAs are the adaptor molecules in protein synthesis 123

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point 124
Major features of the genetic code 125
Messenger RNA contains start and stop signals for protein synthesis 126
The genetic code is nearly universal 126

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons 127
RNA processing generates mature RNA 127
Many exons encode protein domains 128

CHAPTER 5 Exploring Genes and Genomes 135

5.1 The Exploration of Genes Relies on Key Tools 136
Restriction enzymes split DNA into specific fragments 137
Restriction fragments can be separated by gel electrophoresis and visualized 137
DNA can be sequenced by controlled termination of replication 138
DNA probes and genes can be synthesized by automated solid-phase methods 139
Selected DNA sequences can be greatly amplified by the polymerase chain reaction 141
PCR is a powerful technique in medical diagnostics, forensics, and studies of molecular evolution 142
The tools for recombinant DNA technology have been used to identify disease-causing mutations 143

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology 143
Restriction enzymes and DNA ligase are key tools in forming recombinant DNA molecules 143
Plasmids and λ phage are choice vectors for DNA cloning in bacteria 144
Bacterial and yeast artificial chromosomes 147
Specific genes can be cloned from digests of genomic DNA 147
Complementary DNA prepared from mRNA can be expressed in host cells 149
Proteins with new functions can be created through directed changes in DNA 150
Recombinant methods enable the exploration of the functional effects of disease-causing mutations 152

5.3 Complete Genomes Have Been Sequenced and Analyzed 152
The genomes of organisms ranging from bacteria to multicellular eukaryotes have been sequenced 153
The sequence of the human genome has been completed 154
Next-generation sequencing methods enable the rapid determination of a complete genome sequence 155
Comparative genomics has become a powerful research tool 156

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision 157
Gene-expression levels can be comprehensively examined 157
New genes inserted into eukaryotic cells can be efficiently expressed 159
Transgenic animals harbor and express genes introduced into their germ lines 160
Gene disruption and genome editing provide clues to gene function and opportunities for new therapies 160
RNA interference provides an additional tool for disrupting gene expression 162
Tumor-inducing plasmids can be used to introduce new genes into plant cells 163
Human gene therapy holds great promise for medicine 164

CHAPTER 6 Exploring Evolution and Bioinformatics 169

6.1 Homologs Are Descended from a Common Ancestor 170

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology 171
The statistical significance of alignments can be estimated by shuffling 173
Distant evolutionary relationships can be detected through the use of substitution matrices 174
Databases can be searched to identify homologous sequences 177

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships 177
Tertiary structure is more conserved than primary structure 178
Knowledge of three-dimensional structures can aid in the evaluation of sequence alignments 179
Repeated motifs can be detected by aligning sequences with themselves 180
Convergent evolution illustrates common solutions to biochemical challenges 181
Comparison of RNA sequences can be a source of insight into RNA secondary structures 182

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information 183
Horizontal gene transfer events may explain unexpected branches of the evolutionary tree 184

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible 185
Ancient DNA can sometimes be amplified and sequenced 185
Molecular evolution can be examined experimentally 185

CHAPTER 7 Hemoglobin: Portrait of a Protein in Action 191

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme 192
Changes in heme electronic structure upon oxygen binding are the basis for functional imaging studies 193
The structure of myoglobin prevents the release of reactive oxygen species 194
Human hemoglobin is an assembly of four myoglobin-like subunits 195
7.2 Hemoglobin Binds Oxygen Cooperatively 195
Oxygen binding markedly changes the quaternary structure of hemoglobin 197
Hemoglobin cooperativity can be potentially explained by several models 198
Structural changes at the heme groups are transmitted to the $\alpha_1\beta_1$–$\alpha_2\beta_2$ interface 200
2,3-Bisphosphoglycerate in red cells is crucial in determining the oxygen affinity of hemoglobin 200
Carbon monoxide can disrupt oxygen transport by hemoglobin 201
7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen: The Bohr Effect 202
7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease 204
Sickle-cell anemia results from the aggregation of mutated deoxyhemoglobin molecules 205
Thalassemia is caused by an imbalanced production of hemoglobin chains 207
The accumulation of free alpha-hemoglobin chains is prevented 207
Additional globins are encoded in the human genome 208
APPENDIX: Binding Models Can Be Formulated in Quantitative Terms: The Hill Plot and the Concerted Model 210

CHAPTER 8 Enzymes: Basic Concepts and Kinetics 215

8.1 Enzymes are Powerful and Highly Specific Catalysts 216
Many enzymes require cofactors for activity 217
Enzymes can transform energy from one form into another 217
8.2 Gibbs Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes 218
The free-energy change provides information about the spontaneity but not the rate of a reaction 218
The standard free-energy change of a reaction is related to the equilibrium constant 219
Enzymes alter only the reaction rate and not the reaction equilibrium 220
8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State 221
The formation of an enzyme–substrate complex is the first step in enzymatic catalysis 222
The active sites of enzymes have some common features 223
The binding energy between enzyme and substrate is important for catalysis 225
8.4 The Michaelis–Menten Model Accounts for the Kinetic Properties of Many Enzymes 225
Kinetics is the study of reaction rates 225
The steady-state assumption facilitates a description of enzyme kinetics 226
Variations in K_M can have physiological consequences 228
K_M and V_{max} values can be determined by several means 228
K_M and V_{max} values are important enzyme characteristics 229
k_{cat}/K_M is a measure of catalytic efficiency 230
Most biochemical reactions include multiple substrates 231
Allosteric enzymes do not obey Michaelis–Menten kinetics 233
8.5 Enzymes Can Be Inhibited by Specific Molecules 234
The different types of reversible inhibitors are kinetically distinguishable 235
Irreversible inhibitors can be used to map the active site 237
Penicillin irreversibly inactivates a key enzyme in bacterial cell-wall synthesis 239
Transition-state analogs are potent inhibitors of enzymes 240
Catalytic antibodies demonstrate the importance of selective binding of the transition state to enzymatic activity 241
8.6 Enzymes Can Be Studied One Molecule at a Time 242
APPENDIX: Enzymes are Classified on the Basis of the Types of Reactions That They Catalyze 245

CHAPTER 9 Catalytic Strategies 251

A few basic catalytic principles are used by many enzymes 252
9.1 Proteases Facilitate a Fundamentally Difficult Reaction 253
Chymotrypsin possesses a highly reactive serine residue 253
Chymotrypsin action proceeds in two steps linked by a covalently bound intermediate 254
Serine is part of a catalytic triad that also includes histidine and aspartate 255
Catalytic triads are found in other hydrolytic enzymes 258
The catalytic triad has been dissected by site-directed mutagenesis 260
Cysteine, aspartyl, and metalloproteases are other major classes of peptide-cleaving enzymes 260
Protease inhibitors are important drugs 263

9.2 Carbonic Anhydrases Make a Fast Reaction Faster 264

Carbonic anhydrase contains a bound zinc ion essential for catalytic activity 265

Catalysis entails zinc activation of a water molecule 265

A proton shuttle facilitates rapid regeneration of the active form of the enzyme 267

9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions 269

Cleavage is by in-line displacement of 3′-oxygen from phosphorus by magnesium-activated water 269

Restriction enzymes require magnesium for catalytic activity 271

The complete catalytic apparatus is assembled only within complexes of cognate DNA molecules, ensuring specificity 272

Host-cell DNA is protected by the addition of methyl groups to specific bases 274

Type II restriction enzymes have a catalytic core in common and are probably related by horizontal gene transfer 275

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work 275

ATP hydrolysis proceeds by the attack of water on the gamma-phosphoryl group 276

Formation of the transition state for ATP hydrolysis is associated with a substantial conformational change 277

The altered conformation of myosin persists for a substantial period of time 278

Scientists can watch single molecules of myosin move 279

Myosins are a family of enzymes containing P-loop structures 280

CHAPTER 10 Regulatory Strategies 285

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway 286

Allosterically regulated enzymes do not follow Michaelis–Menten kinetics 287

ATCase consists of separable catalytic and regulatory subunits 287

Allosteric interactions in ATCase are mediated by large changes in quaternary structure 288

Allosteric regulators modulate the T-to-R equilibrium 291

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages 292

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity 293

Kinases and phosphatases control the extent of protein phosphorylation 294

Phosphorylation is a highly effective means of regulating the activities of target proteins 296

Cyclic AMP activates protein kinase A by altering the quaternary structure 297

ATP and the target protein bind to a deep cleft in the catalytic subunit of protein kinase A 298

10.4 Many Enzymes Are Activated by Specific Proteolytic Cleavage 299

Chymotrypsinogen is activated by specific cleavage of a single peptide bond 299

Proteolytic activation of chymotrypsinogen leads to the formation of a substrate-binding site 300

The generation of trypsin from trypsinogen leads to the activation of other zymogens 301

Some proteolytic enzymes have specific inhibitors 302

Blood clotting is accomplished by a cascade of zymogen activations 303

Prothrombin requires a vitamin K-dependent modification for activation 304

Fibrinogen is converted by thrombin into a fibrin clot 304

Vitamin K is required for the formation of γ-carboxyglutamate 306

The clotting process must be precisely regulated 307

Hemophilia revealed an early step in clotting 308

CHAPTER 11 Carbohydrates 315

11.1 Monosaccharides Are the Simplest Carbohydrates 316

Many common sugars exist in cyclic forms 318

Pyranose and furanose rings can assume different conformations 320

Glucose is a reducing sugar 321

Monosaccharides are joined to alcohols and amines through glycosidic bonds 322

Phosphorylated sugars are key intermediates in energy generation and biosyntheses 322

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates 323

Sucrose, lactose, and maltose are the common disaccharides 323

Glycogen and starch are storage forms of glucose 324

Cellulose, a structural component of plants, is made of chains of glucose 324

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins 325

Carbohydrates can be linked to proteins through asparagine (*N*-linked) or through serine or threonine (*O*-linked) residues 326

The glycoprotein erythropoietin is a vital hormone 327

Glycosylation functions in nutrient sensing 327

Proteoglycans, composed of polysaccharides and protein, have important structural roles 327

Proteoglycans are important components of cartilage 328

Mucins are glycoprotein components of mucus 329

Protein glycosylation takes place in the lumen of the endoplasmic reticulum and in the Golgi complex 330

Specific enzymes are responsible for oligosaccharide assembly 331
Blood groups are based on protein glycosylation patterns 331
Errors in glycosylation can result in pathological conditions 332
Oligosaccharides can be "sequenced" 332
11.4 Lectins Are Specific Carbohydrate-Binding Proteins 333
Lectins promote interactions between cells 334
Lectins are organized into different classes 334
Influenza virus binds to sialic acid residues 335

CHAPTER 12 Lipids and Cell Membranes 341

Many common features underlie the diversity of biological membranes 342
12.1 Fatty Acids Are Key Constituents of Lipids 342
Fatty acid names are based on their parent hydrocarbons 342
Fatty acids vary in chain length and degree of unsaturation 343
12.2 There Are Three Common Types of Membrane Lipids 344
Phospholipids are the major class of membrane lipids 344
Membrane lipids can include carbohydrate moieties 345
Cholesterol is a lipid based on a steroid nucleus 346
Archaeal membranes are built from ether lipids with branched chains 346
A membrane lipid is an amphipathic molecule containing a hydrophilic and a hydrophobic moiety 347
12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media 348
Lipid vesicles can be formed from phospholipids 348
Lipid bilayers are highly impermeable to ions and most polar molecules 349
12.4 Proteins Carry Out Most Membrane Processes 350
Proteins associate with the lipid bilayer in a variety of ways 351
Proteins interact with membranes in a variety of ways 351
Some proteins associate with membranes through covalently attached hydrophobic groups 354
Transmembrane helices can be accurately predicted from amino acid sequences 354
12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane 356
The fluid mosaic model allows lateral movement but not rotation through the membrane 357
Membrane fluidity is controlled by fatty acid composition and cholesterol content 357
Lipid rafts are highly dynamic complexes formed between cholesterol and specific lipids 358
All biological membranes are asymmetric 358
12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes 359

CHAPTER 13 Membrane Channels and Pumps 367

The expression of transporters largely defines the metabolic activities of a given cell type 368
13.1 The Transport of Molecules Across a Membrane May Be Active or Passive 368
Many molecules require protein transporters to cross membranes 368
Free energy stored in concentration gradients can be quantified 369
13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions and Molecules Across Membranes 370
P-type ATPases couple phosphorylation and conformational changes to pump calcium ions across membranes 370
Digitalis specifically inhibits the Na^+–K^+ pump by blocking its dephosphorylation 373
P-type ATPases are evolutionarily conserved and play a wide range of roles 374
Multidrug resistance highlights a family of membrane pumps with ATP-binding cassette domains 374
13.3 Lactose Permease Is an Archetype of Secondary Transporters That Use One Concentration Gradient to Power the Formation of Another 376
13.4 Specific Channels Can Rapidly Transport Ions Across Membranes 378
Action potentials are mediated by transient changes in Na^+ and K^+ permeability 378
Patch-clamp conductance measurements reveal the activities of single channels 379
The structure of a potassium ion channel is an archetype for many ion-channel structures 379
The structure of the potassium ion channel reveals the basis of ion specificity 380
The structure of the potassium ion channel explains its rapid rate of transport 383
Voltage gating requires substantial conformational changes in specific ion-channel domains 383
A channel can be inactivated by occlusion of the pore: the ball-and-chain model 384
The acetylcholine receptor is an archetype for ligand-gated ion channels 385
Action potentials integrate the activities of several ion channels working in concert 387
Disruption of ion channels by mutations or chemicals can be potentially life-threatening 388
13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells 389
13.6 Specific Channels Increase the Permeability of Some Membranes to Water 390

CHAPTER 14 Signal-Transduction Pathways 397

Signal transduction depends on molecular circuits 398

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves 399

Ligand binding to 7TM receptors leads to the activation of heterotrimeric G proteins 400

Activated G proteins transmit signals by binding to other proteins 402

Cyclic AMP stimulates the phosphorylation of many target proteins by activating protein kinase A 403

G proteins spontaneously reset themselves through GTP hydrolysis 403

Some 7TM receptors activate the phosphoinositide cascade 404

Calcium ion is a widely used second messenger 405

Calcium ion often activates the regulatory protein calmodulin 407

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes 407

The insulin receptor is a dimer that closes around a bound insulin molecule 408

Insulin binding results in the cross-phosphorylation and activation of the insulin receptor 408

The activated insulin-receptor kinase initiates a kinase cascade 409

Insulin signaling is terminated by the action of phosphatases 411

14.3 EGF Signaling: Signal-Transduction Pathways Are Poised to Respond 411

EGF binding results in the dimerization of the EGF receptor 411

The EGF receptor undergoes phosphorylation of its carboxyl-terminal tail 413

EGF signaling leads to the activation of Ras, a small G protein 413

Activated Ras initiates a protein kinase cascade 414

EGF signaling is terminated by protein phosphatases and the intrinsic GTPase activity of Ras 414

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways 415

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases 416

Monoclonal antibodies can be used to inhibit signal-transduction pathways activated in tumors 416

Protein kinase inhibitors can be effective anticancer drugs 417

Cholera and whooping cough are the result of altered G-protein activity 417

Part II TRANSDUCING AND STORING ENERGY

CHAPTER 15 Metabolism: Basic Concepts and Design 423

15.1 Metabolism Is Composed of Many Coupled, Interconnecting Reactions 424

Metabolism consists of energy-yielding and energy-requiring reactions 424

A thermodynamically unfavorable reaction can be driven by a favorable reaction 425

15.2 ATP Is the Universal Currency of Free Energy in Biological Systems 426

ATP hydrolysis is exergonic 426

ATP hydrolysis drives metabolism by shifting the equilibrium of coupled reactions 427

The high phosphoryl potential of ATP results from structural differences between ATP and its hydrolysis products 429

Phosphoryl-transfer potential is an important form of cellular energy transformation 430

15.3 The Oxidation of Carbon Fuels Is an Important Source of Cellular Energy 432

Compounds with high phosphoryl-transfer potential can couple carbon oxidation to ATP synthesis 432

Ion gradients across membranes provide an important form of cellular energy that can be coupled to ATP synthesis 433

Phosphates play a prominent role in biochemical processes 434

Energy from foodstuffs is extracted in three stages 434

15.4 Metabolic Pathways Contain Many Recurring Motifs 435

Activated carriers exemplify the modular design and economy of metabolism 435

Many activated carriers are derived from vitamins 438

Key reactions are reiterated throughout metabolism 440

Metabolic processes are regulated in three principal ways 442

Aspects of metabolism may have evolved from an RNA world 444

CHAPTER 16 Glycolysis and Gluconeogenesis 449

Glucose is generated from dietary carbohydrates 450

Glucose is an important fuel for most organisms 451

16.1 Glycolysis Is an Energy-Conversion Pathway in Many Organisms 451

Hexokinase traps glucose in the cell and begins glycolysis 451

Fructose 1,6-bisphosphate is generated from glucose 6-phosphate 453

The six-carbon sugar is cleaved into two three-carbon fragments 454

Mechanism: Triose phosphate isomerase salvages a three-carbon fragment 455

The oxidation of an aldehyde to an acid powers the formation of a compound with high phosphoryl-transfer potential 457

Mechanism: Phosphorylation is coupled to the oxidation of glyceraldehyde 3-phosphate by a thioester intermediate 458

ATP is formed by phosphoryl transfer from 1,3-bisphosphoglycerate 459

Additional ATP is generated with the formation of pyruvate 460

Two ATP molecules are formed in the conversion of glucose into pyruvate 461

NAD^+ is regenerated from the metabolism of pyruvate 462
Fermentations provide usable energy in the absence of oxygen 464
The binding site for NAD^+ is similar in many dehydrogenases 465
Fructose is converted into glycolytic intermediates by fructokinase 465
Excessive fructose consumption can lead to pathological conditions 466
Galactose is converted into glucose 6-phosphate 466
Many adults are intolerant of milk because they are deficient in lactase 467
Galactose is highly toxic if the transferase is missing 468
16.2 The Glycolytic Pathway Is Tightly Controlled 469
Glycolysis in muscle is regulated to meet the need for ATP 469
The regulation of glycolysis in the liver illustrates the biochemical versatility of the liver 472
A family of transporters enables glucose to enter and leave animal cells 473
Aerobic glycolysis is a property of rapidly growing cells 474
Cancer and endurance training affect glycolysis in a similar fashion 476
16.3 Glucose Can Be Synthesized from Noncarbohydrate Precursors 476
Gluconeogenesis is not a reversal of glycolysis 478
The conversion of pyruvate into phosphoenolpyruvate begins with the formation of oxaloacetate 478
Oxaloacetate is shuttled into the cytoplasm and converted into phosphoenolpyruvate 480
The conversion of fructose 1,6-bisphosphate into fructose 6-phosphate and orthophosphate is an irreversible step 480
The generation of free glucose is an important control point 481
Six high-transfer-potential phosphoryl groups are spent in synthesizing glucose from pyruvate 481
16.4 Gluconeogenesis and Glycolysis Are Reciprocally Regulated 482
Energy charge determines whether glycolysis or gluconeogenesis will be most active 482
The balance between glycolysis and gluconeogenesis in the liver is sensitive to blood-glucose concentration 483
Substrate cycles amplify metabolic signals and produce heat 485
Lactate and alanine formed by contracting muscle are used by other organs 485
Glycolysis and gluconeogenesis are evolutionarily intertwined 487

CHAPTER 17 The Citric Acid Cycle 495

The citric acid cycle harvests high-energy electrons 496
17.1 The Pyruvate Dehydrogenase Complex Links Glycolysis to the Citric Acid Cycle 497
Mechanism: The synthesis of acetyl coenzyme A from pyruvate requires three enzymes and five coenzymes 498
Flexible linkages allow lipoamide to move between different active sites 500
17.2 The Citric Acid Cycle Oxidizes Two-Carbon Units 501
Citrate synthase forms citrate from oxaloacetate and acetyl coenzyme A 502
Mechanism: The mechanism of citrate synthase prevents undesirable reactions 502
Citrate is isomerized into isocitrate 504
Isocitrate is oxidized and decarboxylated to alpha-ketoglutarate 504
Succinyl coenzyme A is formed by the oxidative decarboxylation of alpha-ketoglutarate 505
A compound with high phosphoryl-transfer potential is generated from succinyl coenzyme A 505
Mechanism: Succinyl coenzyme A synthetase transforms types of biochemical energy 506
Oxaloacetate is regenerated by the oxidation of succinate 507
The citric acid cycle produces high-transfer-potential electrons, ATP, and CO_2 508
17.3 Entry to the Citric Acid Cycle and Metabolism Through It Are Controlled 510
The pyruvate dehydrogenase complex is regulated allosterically and by reversible phosphorylation 511
The citric acid cycle is controlled at several points 512
Defects in the citric acid cycle contribute to the development of cancer 513
17.4 The Citric Acid Cycle Is a Source of Biosynthetic Precursors 514
The citric acid cycle must be capable of being rapidly replenished 514
The disruption of pyruvate metabolism is the cause of beriberi and poisoning by mercury and arsenic 515
The citric acid cycle may have evolved from preexisting pathways 516
17.5 The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate 516

CHAPTER 18 Oxidative Phosphorylation 523

18.1 Eukaryotic Oxidative Phosphorylation Takes Place in Mitochondria 524
Mitochondria are bounded by a double membrane 524
Mitochondria are the result of an endosymbiotic event 525
18.2 Oxidative Phosphorylation Depends on Electron Transfer 526
The electron-transfer potential of an electron is measured as redox potential 526
A 1.14-volt potential difference between NADH and molecular oxygen drives electron transport through the chain and favors the formation of a proton gradient 528
18.3 The Respiratory Chain Consists of Four Complexes: Three Proton Pumps and a Physical Link to the Citric Acid Cycle 529
Iron–sulfur clusters are common components of the electron transport chain 531

The high-potential electrons of NADH enter the respiratory chain at NADH-Q oxidoreductase 532
Ubiquinol is the entry point for electrons from $FADH_2$ of flavoproteins 533
Electrons flow from ubiquinol to cytochrome *c* through Q-cytochrome *c* oxidoreductase 533
The Q cycle funnels electrons from a two-electron carrier to a one-electron carrier and pumps protons 535
Cytochrome *c* oxidase catalyzes the reduction of molecular oxygen to water 535
Toxic derivatives of molecular oxygen such as superoxide radicals are scavenged by protective enzymes 538
Electrons can be transferred between groups that are not in contact 540
The conformation of cytochrome *c* has remained essentially constant for more than a billion years 541

18.4 A Proton Gradient Powers the Synthesis of ATP 541

ATP synthase is composed of a proton-conducting unit and a catalytic unit 543
Proton flow through ATP synthase leads to the release of tightly bound ATP: The binding-change mechanism 544
Rotational catalysis is the world's smallest molecular motor 546
Proton flow around the **c** ring powers ATP synthesis 546
ATP synthase and G proteins have several common features 548

18.5 Many Shuttles Allow Movement Across Mitochondrial Membranes 549

Electrons from cytoplasmic NADH enter mitochondria by shuttles 549
The entry of ADP into mitochondria is coupled to the exit of ATP by ATP-ADP translocase 550
Mitochondrial transporters for metabolites have a common tripartite structure 551

18.6 The Regulation of Cellular Respiration Is Governed Primarily by the Need for ATP 552

The complete oxidation of glucose yields about 30 molecules of ATP 552
The rate of oxidative phosphorylation is determined by the need for ATP 553
ATP synthase can be regulated 554
Regulated uncoupling leads to the generation of heat 554
Oxidative phosphorylation can be inhibited at many stages 556
Mitochondrial diseases are being discovered 557
Mitochondria play a key role in apoptosis 557
Power transmission by proton gradients is a central motif of bioenergetics 558

CHAPTER 19 The Light Reactions of Photosynthesis 565

Photosynthesis converts light energy into chemical energy 566

19.1 Photosynthesis Takes Place in Chloroplasts 567

The primary events of photosynthesis take place in thylakoid membranes 567
Chloroplasts arose from an endosymbiotic event 568

19.2 Light Absorption by Chlorophyll Induces Electron Transfer 568

A special pair of chlorophylls initiate charge separation 569
Cyclic electron flow reduces the cytochrome of the reaction center 572

19.3 Two Photosystems Generate a Proton Gradient and NADPH in Oxygenic Photosynthesis 572

Photosystem II transfers electrons from water to plastoquinone and generates a proton gradient 572
Cytochrome *bf* links photosystem II to photosystem I 575
Photosystem I uses light energy to generate reduced ferredoxin, a powerful reductant 575
Ferredoxin–$NADP^+$ reductase converts $NADP^+$ into NADPH 576

19.4 A Proton Gradient across the Thylakoid Membrane Drives ATP Synthesis 578

The ATP synthase of chloroplasts closely resembles those of mitochondria and prokaryotes 578
The activity of chloroplast ATP synthase is regulated 579
Cyclic electron flow through photosystem I leads to the production of ATP instead of NADPH 580
The absorption of eight photons yields one O_2, two NADPH, and three ATP molecules 581

19.5 Accessory Pigments Funnel Energy into Reaction Centers 581

Resonance energy transfer allows energy to move from the site of initial absorbance to the reaction center 582
The components of photosynthesis are highly organized 583
Many herbicides inhibit the light reactions of photosynthesis 584

19.6 The Ability to Convert Light into Chemical Energy Is Ancient 584

Artificial photosynthetic systems may provide clean, renewable energy 585

CHAPTER 20 The Calvin Cycle and the Pentose Phosphate Pathway 589

20.1 The Calvin Cycle Synthesizes Hexoses from Carbon Dioxide and Water 590

Carbon dioxide reacts with ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate 591
Rubisco activity depends on magnesium and carbamate 592
Rubisco activase is essential for rubisco activity 593
Rubisco also catalyzes a wasteful oxygenase reaction: Catalytic imperfection 593
Hexose phosphates are made from phosphoglycerate, and ribulose 1,5-bisphosphate is regenerated 594
Three ATP and two NADPH molecules are used to bring carbon dioxide to the level of a hexose 597
Starch and sucrose are the major carbohydrate stores in plants 597

20.2 The Activity of the Calvin Cycle Depends on Environmental Conditions 598

Rubisco is activated by light-driven changes in proton and magnesium ion concentrations 598

Thioredoxin plays a key role in regulating the Calvin cycle 599

The C_4 pathway of tropical plants accelerates photosynthesis by concentrating carbon dioxide 599

Crassulacean acid metabolism permits growth in arid ecosystems 601

20.3 The Pentose Phosphate Pathway Generates NADPH and Synthesizes Five-Carbon Sugars 601

Two molecules of NADPH are generated in the conversion of glucose 6-phosphate into ribulose 5-phosphate 602

The pentose phosphate pathway and glycolysis are linked by transketolase and transaldolase 602

Mechanism: Transketolase and transaldolase stabilize carbanionic intermediates by different mechanisms 605

20.4 The Metabolism of Glucose 6-Phosphate by the Pentose Phosphate Pathway Is Coordinated with Glycolysis 607

The rate of the pentose phosphate pathway is controlled by the level of $NADP^+$ 607

The flow of glucose 6-phosphate depends on the need for NADPH, ribose 5-phosphate, and ATP 608

The pentose phosphate pathway is required for rapid cell growth 610

Through the looking-glass: The Calvin cycle and the pentose phosphate pathway are mirror images 610

20.5 Glucose 6-Phosphate Dehydrogenase Plays a Key Role in Protection Against Reactive Oxygen Species 610

Glucose 6-phosphate dehydrogenase deficiency causes a drug-induced hemolytic anemia 610

A deficiency of glucose 6-phosphate dehydrogenase confers an evolutionary advantage in some circumstances 612

CHAPTER 21 Glycogen Metabolism 617

Glycogen metabolism is the regulated release and storage of glucose 618

21.1 Glycogen Breakdown Requires the Interplay of Several Enzymes 619

Phosphorylase catalyzes the phosphorolytic cleavage of glycogen to release glucose 1-phosphate 619

Mechanism: Pyridoxal phosphate participates in the phosphorolytic cleavage of glycogen 620

A debranching enzyme also is needed for the breakdown of glycogen 621

Phosphoglucomutase converts glucose 1-phosphate into glucose 6-phosphate 622

The liver contains glucose 6-phosphatase, a hydrolytic enzyme absent from muscle 622

21.2 Phosphorylase Is Regulated by Allosteric Interactions and Reversible Phosphorylation 623

Liver phosphorylase produces glucose for use by other tissues 623

Muscle phosphorylase is regulated by the intracellular energy charge 625

Biochemical characteristics of muscle fiber types differ 625

Phosphorylation promotes the conversion of phosphorylase *b* to phosphorylase *a* 626

Phosphorylase kinase is activated by phosphorylation and calcium ions 626

21.3 Epinephrine and Glucagon Signal the Need for Glycogen Breakdown 627

G proteins transmit the signal for the initiation of glycogen breakdown 627

Glycogen breakdown must be rapidly turned off when necessary 629

The regulation of glycogen phosphorylase became more sophisticated as the enzyme evolved 629

21.4 Glycogen Is Synthesized and Degraded by Different Pathways 630

UDP-glucose is an activated form of glucose 630

Glycogen synthase catalyzes the transfer of glucose from UDP-glucose to a growing chain 630

A branching enzyme forms α-1,6 linkages 631

Glycogen synthase is the key regulatory enzyme in glycogen synthesis 632

Glycogen is an efficient storage form of glucose 632

21.5 Glycogen Breakdown and Synthesis Are Reciprocally Regulated 632

Protein phosphatase 1 reverses the regulatory effects of kinases on glycogen metabolism 633

Insulin stimulates glycogen synthesis by inactivating glycogen synthase kinase 635

Glycogen metabolism in the liver regulates the blood-glucose level 635

A biochemical understanding of glycogen-storage diseases is possible 637

CHAPTER 22 Fatty Acid Metabolism 643

Fatty acid degradation and synthesis mirror each other in their chemical reactions 644

22.1 Triacylglycerols Are Highly Concentrated Energy Stores 645

Dietary lipids are digested by pancreatic lipases 645

Dietary lipids are transported in chylomicrons 646

22.2 The Use of Fatty Acids as Fuel Requires Three Stages of Processing 647

Triacylglycerols are hydrolyzed by hormone-stimulated lipases 647

Free fatty acids and glycerol are released into the blood 648

Fatty acids are linked to coenzyme A before they are oxidized 648

Carnitine carries long-chain activated fatty acids into the mitochondrial matrix 649

Acetyl CoA, NADH, and $FADH_2$ are generated in each round of fatty acid oxidation 650

The complete oxidation of palmitate yields 106 molecules of ATP 652

22.3 Unsaturated and Odd-Chain Fatty Acids Require Additional Steps for Degradation 652

An isomerase and a reductase are required for the oxidation of unsaturated fatty acids 652

Odd-chain fatty acids yield propionyl CoA in the final thiolysis step 654

Vitamin B_{12} contains a corrin ring and a cobalt atom 654

Mechanism: Methylmalonyl CoA mutase catalyzes a rearrangement to form succinyl CoA 655

Fatty acids are also oxidized in peroxisomes 656

Ketone bodies are formed from acetyl CoA when fat breakdown predominates 657

Ketone bodies are a major fuel in some tissues 658

Animals cannot convert fatty acids into glucose 660

Some fatty acids may contribute to the development of pathological conditions 661

22.4 Fatty Acids Are Synthesized by Fatty Acid Synthase 661

Fatty acids are synthesized and degraded by different pathways 661

The formation of malonyl CoA is the committed step in fatty acid synthesis 662

Intermediates in fatty acid synthesis are attached to an acyl carrier protein 662

Fatty acid synthesis consists of a series of condensation, reduction, dehydration, and reduction reactions 662

Fatty acids are synthesized by a multifunctional enzyme complex in animals 664

The synthesis of palmitate requires 8 molecules of acetyl CoA, 14 molecules of NADPH, and 7 molecules of ATP 666

Citrate carries acetyl groups from mitochondria to the cytoplasm for fatty acid synthesis 666

Several sources supply NADPH for fatty acid synthesis 667

Fatty acid metabolism is altered in tumor cells 667

22.5 The Elongation and Unsaturation of Fatty Acids are Accomplished by Accessory Enzyme Systems 668

Membrane-bound enzymes generate unsaturated fatty acids 668

Eicosanoid hormones are derived from polyunsaturated fatty acids 669

Variations on a theme: Polyketide and nonribosomal peptide synthetases resemble fatty acid synthase 670

22.6 Acetyl CoA Carboxylase Plays a Key Role in Controlling Fatty Acid Metabolism 670

Acetyl CoA carboxylase is regulated by conditions in the cell 671

Acetyl CoA carboxylase is regulated by a variety of hormones 671

CHAPTER 23 Protein Turnover and Amino Acid Catabolism 681

23.1 Proteins are Degraded to Amino Acids 682

The digestion of dietary proteins begins in the stomach and is completed in the intestine 682

Cellular proteins are degraded at different rates 682

23.2 Protein Turnover Is Tightly Regulated 683

Ubiquitin tags proteins for destruction 683

The proteasome digests the ubiquitin-tagged proteins 685

The ubiquitin pathway and the proteasome have prokaryotic counterparts 686

Protein degradation can be used to regulate biological function 687

23.3 The First Step in Amino Acid Degradation Is the Removal of Nitrogen 687

Alpha-amino groups are converted into ammonium ions by the oxidative deamination of glutamate 687

Mechanism: Pyridoxal phosphate forms Schiff-base intermediates in aminotransferases 689

Aspartate aminotransferase is an archetypal pyridoxal-dependent transaminase 690

Blood levels of aminotransferases serve a diagnostic function 691

Pyridoxal phosphate enzymes catalyze a wide array of reactions 691

Serine and threonine can be directly deaminated 692

Peripheral tissues transport nitrogen to the liver 692

23.4 Ammonium Ion Is Converted into Urea in Most Terrestrial Vertebrates 693

The urea cycle begins with the formation of carbamoyl phosphate 693

Carbamoyl phosphate synthetase is the key regulatory enzyme for urea synthesis 694

Carbamoyl phosphate reacts with ornithine to begin the urea cycle 694

The urea cycle is linked to gluconeogenesis 696

Urea-cycle enzymes are evolutionarily related to enzymes in other metabolic pathways 696

Inherited defects of the urea cycle cause hyperammonemia and can lead to brain damage 697

Urea is not the only means of disposing of excess nitrogen 698

23.5 Carbon Atoms of Degraded Amino Acids Emerge as Major Metabolic Intermediates 698

Pyruvate is an entry point into metabolism for a number of amino acids 699

Oxaloacetate is an entry point into metabolism for aspartate and asparagine 700

Alpha-ketoglutarate is an entry point into metabolism for five-carbon amino acids 700

Succinyl coenzyme A is a point of entry for several nonpolar amino acids 701

Methionine degradation requires the formation of a key methyl donor, *S*-adenosylmethionine 701

The branched-chain amino acids yield acetyl CoA, acetoacetate, or propionyl CoA 701

Oxygenases are required for the degradation of aromatic amino acids 703

23.6 Inborn Errors of Metabolism Can Disrupt Amino Acid Degradation 705

Phenylketonuria is one of the most common metabolic disorders 706

Determining the basis of the neurological symptoms of phenylketonuria is an active area of research 706

Part III SYNTHESIZING THE MOLECULES OF LIFE

CHAPTER 24 The Biosynthesis of Amino Acids 713

Amino acid synthesis requires solutions to three key biochemical problems 714

24.1 Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia 714

The iron–molybdenum cofactor of nitrogenase binds and reduces atmospheric nitrogen 715

Ammonium ion is assimilated into an amino acid through glutamate and glutamine 717

24.2 Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways 719

Human beings can synthesize some amino acids but must obtain others from their diet 719

Aspartate, alanine, and glutamate are formed by the addition of an amino group to an alpha-ketoacid 720

A common step determines the chirality of all amino acids 721

The formation of asparagine from aspartate requires an adenylated intermediate 721

Glutamate is the precursor of glutamine, proline, and arginine 722

3-Phosphoglycerate is the precursor of serine, cysteine, and glycine 722

Tetrahydrofolate carries activated one-carbon units at several oxidation levels 723

S-Adenosylmethionine is the major donor of methyl groups 724

Cysteine is synthesized from serine and homocysteine 726

High homocysteine levels correlate with vascular disease 726

Shikimate and chorismate are intermediates in the biosynthesis of aromatic amino acids 727

Tryptophan synthase illustrates substrate channeling in enzymatic catalysis 729

24.3 Feedback Inhibition Regulates Amino Acid Biosynthesis 730

Branched pathways require sophisticated regulation 731

The sensitivity of glutamine synthetase to allosteric regulation is altered by covalent modification 732

24.4 Amino Acids Are Precursors of Many Biomolecules 734

Glutathione, a gamma-glutamyl peptide, serves as a sulfhydryl buffer and an antioxidant 734

Nitric oxide, a short-lived signal molecule, is formed from arginine 735

Porphyrins are synthesized from glycine and succinyl coenzyme A 736

Porphyrins accumulate in some inherited disorders of porphyrin metabolism 737

CHAPTER 25 Nucleotide Biosynthesis 743

Nucleotides can be synthesized by de novo or salvage pathways 744

25.1 The Pyrimidine Ring Is Assembled de Novo or Recovered by Salvage Pathways 744

Bicarbonate and other oxygenated carbon compounds are activated by phosphorylation 745

The side chain of glutamine can be hydrolyzed to generate ammonia 745

Intermediates can move between active sites by channeling 745

Orotate acquires a ribose ring from PRPP to form a pyrimidine nucleotide and is converted into uridylate 746

Nucleotide mono-, di-, and triphosphates are interconvertible 747

CTP is formed by amination of UTP 747

Salvage pathways recycle pyrimidine bases 748

25.2 Purine Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways 748

The purine ring system is assembled on ribose phosphate 749

The purine ring is assembled by successive steps of activation by phosphorylation followed by displacement 749

AMP and GMP are formed from IMP 751

Enzymes of the purine synthesis pathway associate with one another in vivo 752

Salvage pathways economize intracellular energy expenditure 752

25.3 Deoxyribonucleotides Are Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism 753

Mechanism: A tyrosyl radical is critical to the action of ribonucleotide reductase 753

Stable radicals other than tyrosyl radical are employed by other ribonucleotide reductases 755

Thymidylate is formed by the methylation of deoxyuridylate 755

Dihydrofolate reductase catalyzes the regeneration of tetrahydrofolate, a one-carbon carrier 756

Several valuable anticancer drugs block the synthesis of thymidylate 757

25.4 Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition 758

Pyrimidine biosynthesis is regulated by aspartate transcarbamoylase 758

The synthesis of purine nucleotides is controlled by feedback inhibition at several sites 758

The synthesis of deoxyribonucleotides is controlled by the regulation of ribonucleotide reductase 759

25.5 Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions 760

The loss of adenosine deaminase activity results in severe combined immunodeficiency 760

Gout is induced by high serum levels of urate 761

Lesch–Nyhan syndrome is a dramatic consequence of mutations in a salvage-pathway enzyme 761

Folic acid deficiency promotes birth defects such as spina bifida 762

CHAPTER 26 The Biosynthesis of Membrane Lipids and Steroids 767

26.1 Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols 768

The synthesis of phospholipids requires an activated intermediate 769

Some phospholipids are synthesized from an activated alcohol 770

Phosphatidylcholine is an abundant phospholipid 770

Excess choline is implicated in the development of heart disease 771

Base-exchange reactions can generate phospholipids 771

Sphingolipids are synthesized from ceramide 772

Gangliosides are carbohydrate-rich sphingolipids that contain acidic sugars 772

Sphingolipids confer diversity on lipid structure and function 773

Respiratory distress syndrome and Tay–Sachs disease result from the disruption of lipid metabolism 774

Ceramide metabolism stimulates tumor growth 774

Phosphatidic acid phosphatase is a key regulatory enzyme in lipid metabolism 775

26.2 Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages 776

The synthesis of mevalonate, which is activated as isopentenyl pyrophosphate, initiates the synthesis of cholesterol 776

Squalene (C_{30}) is synthesized from six molecules of isopentenyl pyrophosphate (C_5) 777

Squalene cyclizes to form cholesterol 778

26.3 The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels 779

Lipoproteins transport cholesterol and triacylglycerols throughout the organism 782

Low-density lipoproteins play a central role in cholesterol metabolism 784

The absence of the LDL receptor leads to hypercholesterolemia and atherosclerosis 784

Mutations in the LDL receptor prevent LDL release and result in receptor destruction 785

Cycling of the LDL receptor is regulated 787

HDL appears to protect against atherosclerosis 787

The clinical management of cholesterol levels can be understood at a biochemical level 788

26.4 Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones 788

Letters identify the steroid rings and numbers identify the carbon atoms 790

Steroids are hydroxylated by cytochrome P450 monooxygenases that use NADPH and O_2 790

The cytochrome P450 system is widespread and performs a protective function 791

Pregnenolone, a precursor of many other steroids, is formed from cholesterol by cleavage of its side chain 792

Progesterone and corticosteroids are synthesized from pregnenolone 792

Androgens and estrogens are synthesized from pregnenolone 792

Vitamin D is derived from cholesterol by the ring-splitting activity of light 794

CHAPTER 27 The Integration of Metabolism 801

27.1 Caloric Homeostasis Is a Means of Regulating Body Weight 802

27.2 The Brain Plays a Key Role in Caloric Homeostasis 804

Signals from the gastrointestinal tract induce feelings of satiety 804

Leptin and insulin regulate long-term control over caloric homeostasis 805

Leptin is one of several hormones secreted by adipose tissue 806

Leptin resistance may be a contributing factor to obesity 806

Dieting is used to combat obesity 807

27.3 Diabetes Is a Common Metabolic Disease Often Resulting from Obesity 807

Insulin initiates a complex signal-transduction pathway in muscle 808

Metabolic syndrome often precedes type 2 diabetes 809

Excess fatty acids in muscle modify metabolism 810

Insulin resistance in muscle facilitates pancreatic failure 810

Metabolic derangements in type 1 diabetes result from insulin insufficiency and glucagon excess 812

27.4 Exercise Beneficially Alters the Biochemistry of Cells 813

Mitochondrial biogenesis is stimulated by muscular activity 813

Fuel choice during exercise is determined by the intensity and duration of activity 813

27.5 Food Intake and Starvation Induce Metabolic Changes 816

The starved–fed cycle is the physiological response to a fast 816

Metabolic adaptations in prolonged starvation minimize protein degradation 818

27.6 Ethanol Alters Energy Metabolism in the Liver 819

Ethanol metabolism leads to an excess of NADH 820

Excess ethanol consumption disrupts vitamin metabolism 821

CHAPTER 28 DNA Replication, Repair, and Recombination 827

28.1 DNA Replication Proceeds by the Polymerization of Deoxyribonucleoside Triphosphates Along a Template 828

DNA polymerases require a template and a primer 829
All DNA polymerases have structural features in common 829
Two bound metal ions participate in the polymerase reaction 829
The specificity of replication is dictated by complementarity of shape between bases 830
An RNA primer synthesized by primase enables DNA synthesis to begin 831
One strand of DNA is made continuously, whereas the other strand is synthesized in fragments 831
DNA ligase joins ends of DNA in duplex regions 832
The separation of DNA strands requires specific helicases and ATP hydrolysis 832

28.2 DNA Unwinding and Supercoiling Are Controlled by Topoisomerases 833
The linking number of DNA, a topological property, determines the degree of supercoiling 835
Topoisomerases prepare the double helix for unwinding 836
Type I topoisomerases relax supercoiled structures 836
Type II topoisomerases can introduce negative supercoils through coupling to ATP hydrolysis 837

28.3 DNA Replication Is Highly Coordinated 839
DNA replication requires highly processive polymerases 839
The leading and lagging strands are synthesized in a coordinated fashion 840
DNA replication in *Escherichia coli* begins at a unique site 842
DNA synthesis in eukaryotes is initiated at multiple sites 843
Telomeres are unique structures at the ends of linear chromosomes 844
Telomeres are replicated by telomerase, a specialized polymerase that carries its own RNA template 845

28.4 Many Types of DNA Damage Can Be Repaired 845
Errors can arise in DNA replication 846
Bases can be damaged by oxidizing agents, alkylating agents, and light 846
DNA damage can be detected and repaired by a variety of systems 847
The presence of thymine instead of uracil in DNA permits the repair of deaminated cytosine 849
Some genetic diseases are caused by the expansion of repeats of three nucleotides 850
Many cancers are caused by the defective repair of DNA 850
Many potential carcinogens can be detected by their mutagenic action on bacteria 852

28.5 DNA Recombination Plays Important Roles in Replication, Repair, and Other Processes 852
RecA can initiate recombination by promoting strand invasion 853
Some recombination reactions proceed through Holliday-junction intermediates 854

CHAPTER 29 RNA Synthesis and Processing 859
RNA synthesis comprises three stages: Initiation, elongation, and termination 860

29.1 RNA Polymerases Catalyze Transcription 861
RNA chains are formed de novo and grow in the 5′-to-3′ direction 862
RNA polymerases backtrack and correct errors 863
RNA polymerase binds to promoter sites on the DNA template to initiate transcription 864
Sigma subunits of RNA polymerase recognize promoter sites 865
RNA polymerases must unwind the template double helix for transcription to take place 865
Elongation takes place at transcription bubbles that move along the DNA template 866
Sequences within the newly transcribed RNA signal termination 866
Some messenger RNAs directly sense metabolite concentrations 867
The *rho* protein helps to terminate the transcription of some genes 868
Some antibiotics inhibit transcription 869
Precursors of transfer and ribosomal RNA are cleaved and chemically modified after transcription in prokaryotes 870

29.2 Transcription in Eukaryotes Is Highly Regulated 871
Three types of RNA polymerase synthesize RNA in eukaryotic cells 872
Three common elements can be found in the RNA polymerase II promoter region 874
The TFIID protein complex initiates the assembly of the active transcription complex 874
Multiple transcription factors interact with eukaryotic promoters 875
Enhancer sequences can stimulate transcription at start sites thousands of bases away 876

29.3 The Transcription Products of Eukaryotic Polymerases Are Processed 876
RNA polymerase I produces three ribosomal RNAs 877
RNA polymerase III produces transfer RNA 877
The product of RNA polymerase II, the pre-mRNA transcript, acquires a 5′ cap and a 3′ poly(A) tail 878
Small regulatory RNAs are cleaved from larger precursors 879
RNA editing changes the proteins encoded by mRNA 879
Sequences at the ends of introns specify splice sites in mRNA precursors 880
Splicing consists of two sequential transesterification reactions 881
Small nuclear RNAs in spliceosomes catalyze the splicing of mRNA precursors 882
Transcription and processing of mRNA are coupled 883
Mutations that affect pre-mRNA splicing cause disease 884
Most human pre-mRNAS can be spliced in alternative ways to yield different proteins 885

29.4 The Discovery of Catalytic RNA was Revealing in Regard to Both Mechanism and Evolution 886

CHAPTER 30 Protein Synthesis 893

30.1 Protein Synthesis Requires the Translation of Nucleotide Sequences into Amino Acid Sequences 894

- The synthesis of long proteins requires a low error frequency 894
- Transfer RNA molecules have a common design 895
- Some transfer RNA molecules recognize more than one codon because of wobble in base-pairing 897

30.2 Aminoacyl Transfer RNA Synthetases Read the Genetic Code 898

- Amino acids are first activated by adenylation 898
- Aminoacyl-tRNA synthetases have highly discriminating amino acid activation sites 899
- Proofreading by aminoacyl-tRNA synthetases increases the fidelity of protein synthesis 900
- Synthetases recognize various features of transfer RNA molecules 901
- Aminoacyl-tRNA synthetases can be divided into two classes 901

30.3 The Ribosome Is the Site of Protein Synthesis 902

- Ribosomal RNAs (5S, 16S, and 23S rRNA) play a central role in protein synthesis 903
- Ribosomes have three tRNA-binding sites that bridge the 30s and 50s subunits 905
- The start signal is usually AUG preceded by several bases that pair with 16S rRNA 905
- Bacterial protein synthesis is initiated by formylmethionyl transfer RNA 906
- Formylmethionyl-tRNA$_f$ is placed in the P site of the ribosome in the formation of the 70S initiation complex 907
- Elongation factors deliver aminoacyl-tRNA to the ribosome 907
- Peptidyl transferase catalyzes peptide-bond synthesis 908
- The formation of a peptide bond is followed by the GTP-driven translocation of tRNAs and mRNA 909
- Protein synthesis is terminated by release factors that read stop codons 910

30.4 Eukaryotic Protein Synthesis Differs from Bacterial Protein Synthesis Primarily in Translation Initiation 911

- Mutations in initiation factor 2 cause a curious pathological condition 913

30.5 A Variety of Antibiotics and Toxins Can Inhibit Protein Synthesis 913

- Some antibiotics inhibit protein synthesis 914
- Diphtheria toxin blocks protein synthesis in eukaryotes by inhibiting translocation 914
- Ricin fatally modifies 28S ribosomal RNA 915

30.6 Ribosomes Bound to the Endoplasmic Reticulum Manufacture Secretory and Membrane Proteins 915

- Protein synthesis begins on ribosomes that are free in the cytoplasm 916
- Signal sequences mark proteins for translocation across the endoplasmic reticulum membrane 916
- Transport vesicles carry cargo proteins to their final destination 918

CHAPTER 31 The Control of Gene Expression in Prokaryotes 925

31.1 Many DNA-Binding Proteins Recognize Specific DNA Sequences 926

- The helix-turn-helix motif is common to many prokaryotic DNA-binding proteins 927

31.2 Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons 927

- An operon consists of regulatory elements and protein-encoding genes 928
- The *lac* repressor protein in the absence of lactose binds to the operator and blocks transcription 929
- Ligand binding can induce structural changes in regulatory proteins 930
- The operon is a common regulatory unit in prokaryotes 930
- Transcription can be stimulated by proteins that contact RNA polymerase 931

31.3 Regulatory Circuits Can Result in Switching Between Patterns of Gene Expression 932

- The λ repressor regulates its own expression 932
- A circuit based on the λ repressor and Cro forms a genetic switch 933
- Many prokaryotic cells release chemical signals that regulate gene expression in other cells 933
- Biofilms are complex communities of prokaryotes 934

31.4 Gene Expression Can Be Controlled at Posttranscriptional Levels 935

- Attenuation is a prokaryotic mechanism for regulating transcription through the modulation of nascent RNA secondary structure 935

CHAPTER 32 The Control of Gene Expression in Eukaryotes 941

32.1 Eukaryotic DNA Is Organized into Chromatin 943

- Nucleosomes are complexes of DNA and histones 943
- DNA wraps around histone octamers to form nucleosomes 943

32.2 Transcription Factors Bind DNA and Regulate Transcription Initiation 945

- A range of DNA-binding structures are employed by eukaryotic DNA-binding proteins 945
- Activation domains interact with other proteins 946
- Multiple transcription factors interact with eukaryotic regulatory regions 946
- Enhancers can stimulate transcription in specific cell types 946
- Induced pluripotent stem cells can be generated by introducing four transcription factors into differentiated cells 947

32.3 The Control of Gene Expression Can Require Chromatin Remodeling 948

The methylation of DNA can alter patterns of gene expression 949

Steroids and related hydrophobic molecules pass through membranes and bind to DNA-binding receptors 949

Nuclear hormone receptors regulate transcription by recruiting coactivators to the transcription complex 950

Steroid-hormone receptors are targets for drugs 951

Chromatin structure is modulated through covalent modifications of histone tails 952

Histone deacetylases contribute to transcriptional repression 953

32.4 Eukaryotic Gene Expression Can Be Controlled at Posttranscriptional Levels 954

Genes associated with iron metabolism are translationally regulated in animals 954

Small RNAs regulate the expression of many eukaryotic genes 956

Part IV RESPONDING TO ENVIRONMENTAL CHANGES

CHAPTER 33 Sensory Systems 961

33.1 A Wide Variety of Organic Compounds Are Detected by Olfaction 962

Olfaction is mediated by an enormous family of seven-transmembrane-helix receptors 962

Odorants are decoded by a combinatorial mechanism 964

33.2 Taste Is a Combination of Senses That Function by Different Mechanisms 966

Sequencing of the human genome led to the discovery of a large family of 7TM bitter receptors 967

A heterodimeric 7TM receptor responds to sweet compounds 968

Umami, the taste of glutamate and aspartate, is mediated by a heterodimeric receptor related to the sweet receptor 969

Salty tastes are detected primarily by the passage of sodium ions through channels 969

Sour tastes arise from the effects of hydrogen ions (acids) on channels 969

33.3 Photoreceptor Molecules in the Eye Detect Visible Light 970

Rhodopsin, a specialized 7TM receptor, absorbs visible light 970

Light absorption induces a specific isomerization of bound 11-*cis*-retinal 971

Light-induced lowering of the calcium level coordinates recovery 972

Color vision is mediated by three cone receptors that are homologs of rhodopsin 973

Rearrangements in the genes for the green and red pigments lead to “color blindness” 974

33.4 Hearing Depends on the Speedy Detection of Mechanical Stimuli 975

Hair cells use a connected bundle of stereocilia to detect tiny motions 975

Mechanosensory channels have been identified in *Drosophila* and vertebrates 976

33.5 Touch Includes the Sensing of Pressure, Temperature, and Other Factors 977

Studies of capsaicin reveal a receptor for sensing high temperatures and other painful stimuli 977

CHAPTER 34 The Immune System 981

Innate immunity is an evolutionarily ancient defense system 982

The adaptive immune system responds by using the principles of evolution 984

34.1 Antibodies Possess Distinct Antigen-Binding and Effector Units 985

34.2 Antibodies Bind Specific Molecules Through Hypervariable Loops 988

The immunoglobulin fold consists of a beta-sandwich framework with hypervariable loops 988

X-ray analyses have revealed how antibodies bind antigens 989

Large antigens bind antibodies with numerous interactions 990

34.3 Diversity Is Generated by Gene Rearrangements 991

J (joining) genes and D (diversity) genes increase antibody diversity 991

More than 10^8 antibodies can be formed by combinatorial association and somatic mutation 992

The oligomerization of antibodies expressed on the surfaces of immature B cells triggers antibody secretion 993

Different classes of antibodies are formed by the hopping of V_H genes 994

34.4 Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors 995

Peptides presented by MHC proteins occupy a deep groove flanked by alpha helices 996

T-cell receptors are antibody-like proteins containing variable and constant regions 998

CD8 on cytotoxic T cells acts in concert with T-cell receptors 998

Helper T cells stimulate cells that display foreign peptides bound to class II MHC proteins 1000

Helper T cells rely on the T-cell receptor and CD4 to recognize foreign peptides on antigen-presenting cells 1000

MHC proteins are highly diverse 1002

Human immunodeficiency viruses subvert the immune system by destroying helper T cells 1003

34.5 The Immune System Contributes to the Prevention and the Development of Human Diseases 1004

T cells are subjected to positive and negative selection in the thymus 1004

Autoimmune diseases result from the generation of immune responses against self-antigens 1005

The immune system plays a role in cancer prevention 1005
Vaccines are a powerful means to prevent and eradicate disease 1006

CHAPTER 35 Molecular Motors 1011

35.1 Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily 1012
Molecular motors are generally oligomeric proteins with an ATPase core and an extended structure 1012
ATP binding and hydrolysis induce changes in the conformation and binding affinity of motor proteins 1014
35.2 Myosins Move Along Actin Filaments 1016
Actin is a polar, self-assembling, dynamic polymer 1016
Myosin head domains bind to actin filaments 1018
Motions of single motor proteins can be directly observed 1018
Phosphate release triggers the myosin power stroke 1019
Muscle is a complex of myosin and actin 1019
The length of the lever arm determines motor velocity 1022
35.3 Kinesin and Dynein Move Along Microtubules 1022
Microtubules are hollow cylindrical polymers 1022
Kinesin motion is highly processive 1024
35.4 A Rotary Motor Drives Bacterial Motion 1026
Bacteria swim by rotating their flagella 1026
Proton flow drives bacterial flagellar rotation 1026
Bacterial chemotaxis depends on reversal of the direction of flagellar rotation 1028

CHAPTER 36 Drug Development 1033

36.1 The Development of Drugs Presents Huge Challenges 1034
Drug candidates must be potent and selective modulators of their targets 1035
Drugs must have suitable properties to reach their targets 1036
Toxicity can limit drug effectiveness 1040
36.2 Drug Candidates Can Be Discovered by Serendipity, Screening, or Design 1041
Serendipitous observations can drive drug development 1041
Natural products are a valuable source of drugs and drug leads 1043
Screening libraries of synthetic compounds expands the opportunity for identification of drug leads 1044
Drugs can be designed on the basis of three-dimensional structural information about their targets 1046
36.3 Analyses of Genomes Hold Great Promise for Drug Discovery 1048
Potential targets can be identified in the human proteome 1048
Animal models can be developed to test the validity of potential drug targets 1049
Potential targets can be identified in the genomes of pathogens 1050
Genetic differences influence individual responses to drugs 1050
36.4 The Clinical Development of Drugs Proceeds Through Several Phases 1051
Clinical trials are time consuming and expensive 1052
The evolution of drug resistance can limit the utility of drugs for infectious agents and cancer 1053

Answers to Problems A1

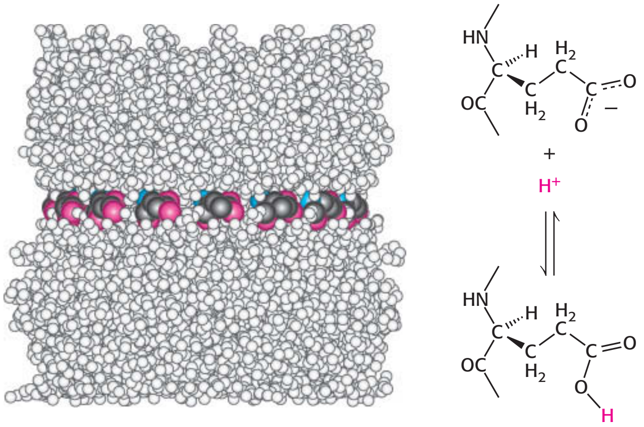
Selected Readings B1

Index C1

CHAPTER

1

Biochemistry: An Evolving Science



Chemistry in action. Human activities require energy. The interconversion of different forms of energy requires large biochemical machines comprising many thousands of atoms such as the complex shown above. Yet, the functions of these elaborate assemblies depend on simple chemical processes such as the protonation and deprotonation of the carboxylic acid groups shown on the right. The photograph is of Nobel Prize winners Peter Agre, M.D., and Carol Greider, Ph.D., who used, respectively, biochemical techniques to reveal key mechanisms of how water is transported into and out of cells, and how chromosomes are replicated faithfully. [Keith Weller for Johns Hopkins Medicine.]

Biochemistry is the study of the chemistry of life processes. Since the discovery that biological molecules such as urea could be synthesized from nonliving components in 1828, scientists have explored the chemistry of life with great intensity. Through these investigations, many of the most fundamental mysteries of how living things function at a biochemical level have now been solved. However, much remains to be investigated. As is often the case, each discovery raises at least as many new questions as it answers. Furthermore, we are now in an age of unprecedented opportunity for the application of our tremendous knowledge of biochemistry to problems in medicine, dentistry, agriculture, forensics, anthropology, environmental sciences, alternative energy, and many other fields. We begin our journey into biochemistry with one of the most startling discoveries of the past century: namely, the great unity of all living things at the biochemical level.

OUTLINE

1.1 Biochemical Unity Underlies Biological Diversity

1.2 DNA Illustrates the Interplay Between Form and Function

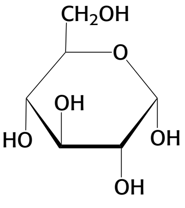
1.3 Concepts from Chemistry Explain the Properties of Biological Molecules

1.4 The Genomic Revolution Is Transforming Biochemistry, Medicine, and Other Fields

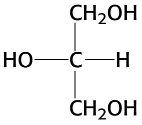
1.1 Biochemical Unity Underlies Biological Diversity

The biological world is magnificently diverse. The animal kingdom is rich with species ranging from nearly microscopic insects to elephants and whales. The plant kingdom includes species as small and relatively simple as algae and as large and complex as giant sequoias. This diversity extends further when we descend into the microscopic world. Organisms such as protozoa, yeast, and bacteria are present with great diversity in water, in soil, and on or within larger organisms. Some organisms can survive and even thrive in seemingly hostile environments such as hot springs and glaciers.

The development of the microscope revealed a key unifying feature that underlies this diversity. Large organisms are built up of *cells,* resembling, to some extent, single-celled microscopic organisms. The construction of animals, plants, and microorganisms from cells suggested that these diverse organisms might have more in common than is apparent from their outward appearance. With the development of biochemistry, this suggestion has been tremendously supported and expanded. At the biochemical level, all organisms have many common features (Figure 1.1).



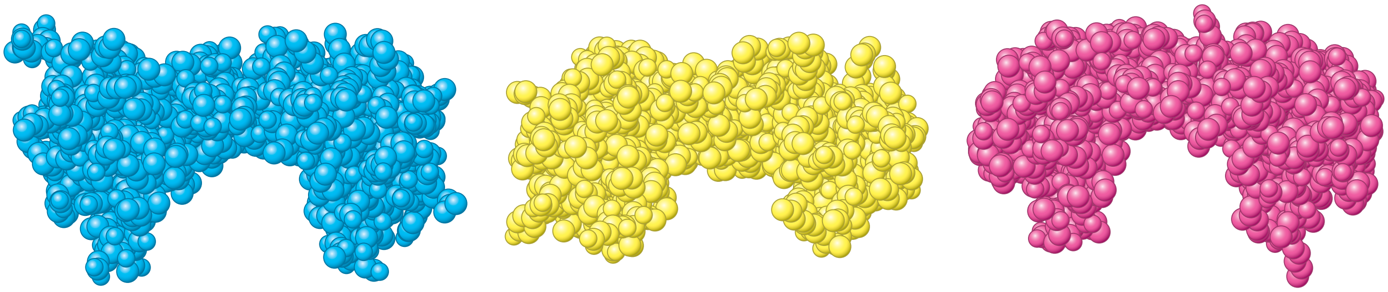
Glucose



Glycerol

As mentioned earlier, biochemistry is the study of the chemistry of life processes. These processes entail the interplay of two different classes of molecules: large molecules such as proteins and nucleic acids, referred to as *biological macromolecules,* and low-molecular-weight molecules such as glucose and glycerol, referred to as *metabolites,* that are chemically transformed in biological processes.

Members of both these classes of molecules are common, with minor variations, to all living things. For example, *deoxyribonucleic acid* (DNA) stores genetic information in all cellular organisms. *Proteins,* the macromolecules that are key participants in most biological processes, are built from the same set of 20 building blocks in all organisms. Furthermore, proteins that play similar roles in different organisms often have very similar three-dimensional structures (Figure 1.1).



Sulfolobus archaea *Arabidopsis thaliana* *Homo sapiens*

FIGURE 1.1 **Biological diversity and similarity.** The shape of a key molecule in gene regulation (the TATA-box-binding protein) is similar in three very different organisms that are separated from one another by billions of years of evolution. [(Left) Eye of Science/Science Source; (middle) Holt Studios/Photo Researchers; (right) Time Life Pictures/Getty Images.]

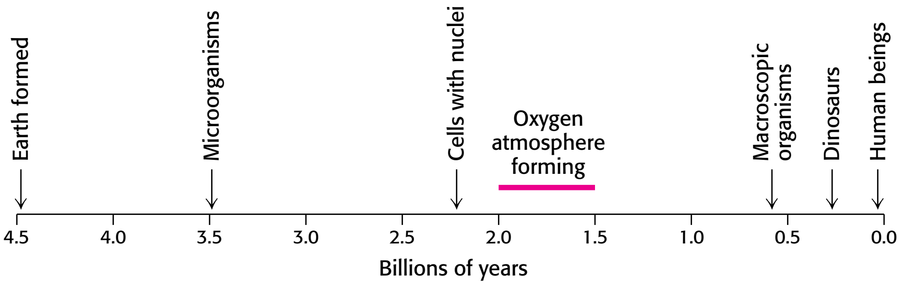


FIGURE 1.2 A possible time line for biochemical evolution. Selected key events are indicated. Note that life on Earth began approximately 3.5 billion years ago, whereas human beings emerged quite recently.

Key metabolic processes also are common to many organisms. For example, the set of chemical transformations that converts glucose and oxygen into carbon dioxide and water is essentially identical in simple bacteria such as *Escherichia coli (E. coli)* and human beings. Even processes that appear to be quite distinct often have common features at the biochemical level. Remarkably, the biochemical processes by which plants capture light energy and convert it into more-useful forms are strikingly similar to steps used in animals to capture energy released from the breakdown of glucose.

These observations overwhelmingly suggest that all living things on Earth have a common ancestor and that modern organisms have evolved from this ancestor into their present forms. Geological and biochemical findings support a time line for this evolutionary path (Figure 1.2). On the basis of their biochemical characteristics, the diverse organisms of the modern world can be divided into three fundamental groups called *domains: Eukarya* (eukaryotes), *Bacteria,* and *Archaea.* Domain Eukarya comprises all multicellular organisms, including human beings as well as many microscopic unicellular organisms such as yeast. The defining characteristic of *eukaryotes* is the presence of a well-defined nucleus within each cell. Unicellular organisms such as bacteria, which lack a nucleus, are referred to as *prokaryotes.* The prokaryotes were reclassified as two separate domains in response to Carl Woese's discovery in 1977 that certain bacteria-like organisms are biochemically quite distinct from other previously characterized bacterial species. These organisms, now recognized as having diverged from bacteria early in evolution, are the *archaea.* Evolutionary paths from a common ancestor to modern organisms can be deduced on the basis of biochemical information. One such path is shown in Figure 1.3.

Much of this book will explore the chemical reactions and the associated biological macromolecules and metabolites that are found in biological processes common to all organisms. The unity of life at the biochemical level makes this approach possible. At the same time, different organisms have specific needs, depending on the particular biological niche in which they evolved and live. By comparing and contrasting details of particular biochemical pathways in different organisms, we can learn how biological challenges are solved at the biochemical level. In most cases, these challenges are addressed by the adaptation of existing macromolecules to new roles rather than by the evolution of entirely new ones.

Biochemistry has been greatly enriched by our ability to examine the three-dimensional structures of biological macromolecules in great detail. Some of these structures

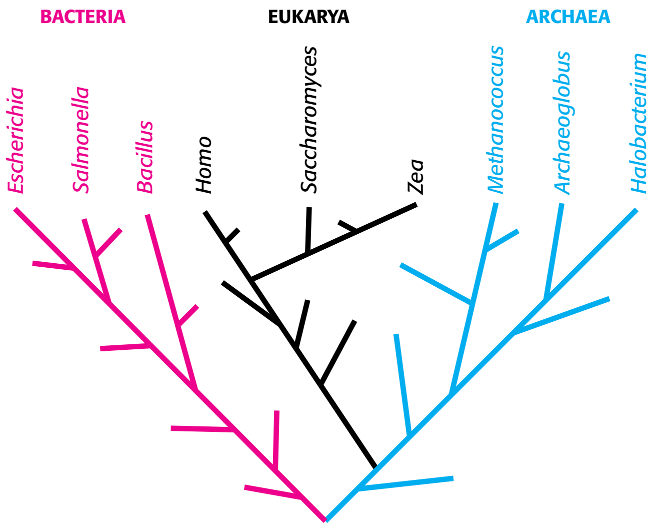


FIGURE 1.3 The tree of life. A possible evolutionary path from a common ancestor approximately 3.5 billion years ago at the bottom of the tree to organisms found in the modern world at the top.

are simple and elegant, whereas others are incredibly complicated. In any case, these structures provide an essential framework for understanding function. We begin our exploration of the interplay between structure and function with the genetic material, DNA.

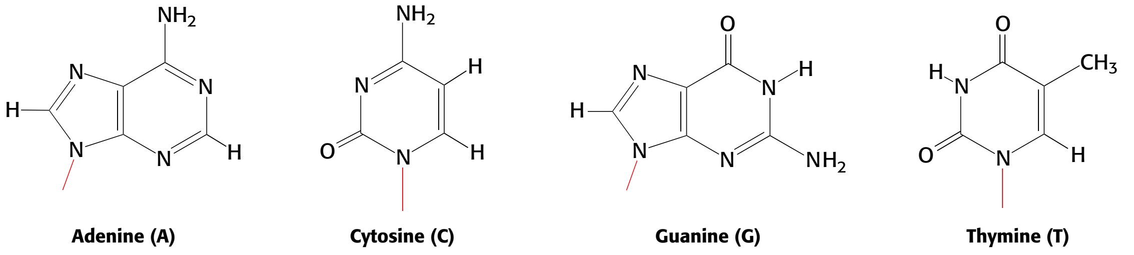
1.2 DNA Illustrates the Interplay Between Form and Function

A fundamental biochemical feature common to all cellular organisms is the use of DNA for the storage of genetic information. The discovery that DNA plays this central role was first made in studies of bacteria in the 1940s. This discovery was followed by a compelling proposal for the three-dimensional structure of DNA in 1953, an event that set the stage for many of the advances in biochemistry and many other fields, extending to the present.

The structure of DNA powerfully illustrates a basic principle common to all biological macromolecules: the intimate relation between structure and function. The remarkable properties of this chemical substance allow it to function as a very efficient and robust vehicle for storing information. We start with an examination of the covalent structure of DNA and its extension into three dimensions.

DNA is constructed from four building blocks

DNA is a *linear polymer* made up of four different types of monomers. It has a fixed backbone from which protrude variable substituents, referred to as bases (Figure 1.4). The backbone is built of repeating sugar–phosphate units. The sugars are molecules of *deoxyribose* from which DNA receives its name. Each sugar is connected to two phosphate groups through different linkages. Moreover, each sugar is oriented in the same way, and so each DNA strand has directionality, with one end distinguishable from the other. Joined to each deoxyribose is one of four possible bases: adenine (A), cytosine (C), guanine (G), and thymine (T).



These bases are connected to the sugar components in the DNA backbone through the bonds shown in black in Figure 1.4. All four bases are planar but differ significantly in other respects. Thus, each monomer of DNA consists of a sugar–phosphate unit and one of four bases attached to the sugar. These bases can be arranged in any order along a strand of DNA.

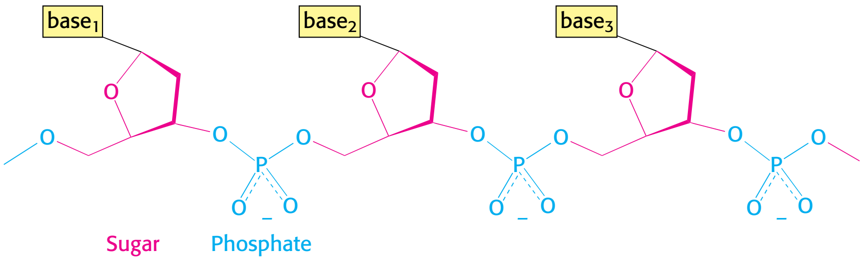


FIGURE 1.4 Covalent structure of DNA. Each unit of the polymeric structure is composed of a sugar (deoxyribose), a phosphate, and a variable base that protrudes from the sugar–phosphate backbone.

Two single strands of DNA combine to form a double helix

Most DNA molecules consist of not one but two strands (Figure 1.5). In 1953, James Watson and Francis Crick deduced the arrangement of these strands and proposed a three-dimensional structure for DNA molecules. This structure is a *double helix* composed of two intertwined strands arranged such that the sugar–phosphate backbone lies on the outside and the bases on the inside. The key to this structure is that the bases form *specific base pairs* (bp) held together by *hydrogen bonds* (Section 1.3): adenine pairs with thymine (A–T) and guanine pairs with cytosine (G–C), as shown in Figure 1.6. Hydrogen bonds are much weaker than *covalent bonds* such as the carbon–carbon or carbon–nitrogen bonds that define the structures of the bases themselves. Such weak bonds are crucial to biochemical systems; they are weak enough to be reversibly broken in biochemical processes, yet they are strong enough, particularly when many form simultaneously, to help stabilize specific structures such as the double helix.

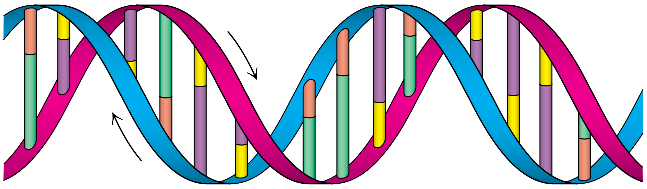


FIGURE 1.5 The double helix. The double-helical structure of DNA proposed by Watson and Crick. The sugar–phosphate backbones of the two chains are shown in red and blue, and the bases are shown in green, purple, orange, and yellow. The two strands are antiparallel, running in opposite directions with respect to the axis of the double helix, as indicated by the arrows.

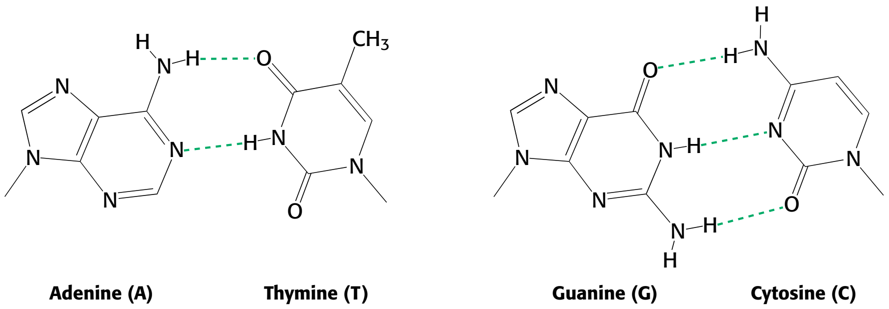


FIGURE 1.6 Watson–Crick base pairs. Adenine pairs with thymine (A–T), and guanine with cytosine (G–C). The dashed green lines represent hydrogen bonds.

DNA structure explains heredity and the storage of information

The structure proposed by Watson and Crick has two properties of central importance to the role of DNA as the hereditary material. First, the structure is compatible with any sequence of bases. While the bases are distinct in structure, the base pairs have essentially the same shape (Figure 1.6) and thus fit equally well into the center of the double-helical structure of any sequence. Without any constraints, the sequence of bases along a DNA strand can act as an efficient means of storing information. Indeed, the sequence of bases along DNA strands is how genetic information is stored. The DNA sequence determines the sequences of the ribonucleic acid (RNA) and protein molecules that carry out most of the activities within cells.

Second, because of base-pairing, the sequence of bases along one strand completely determines the sequence along the other strand. As Watson and Crick so coyly wrote: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." Thus, if the DNA double helix is separated into two single strands, each strand can act as a template for the generation of its partner strand through specific base-pair formation (Figure 1.7). The three-dimensional structure of DNA beautifully illustrates the close connection between molecular form and function.

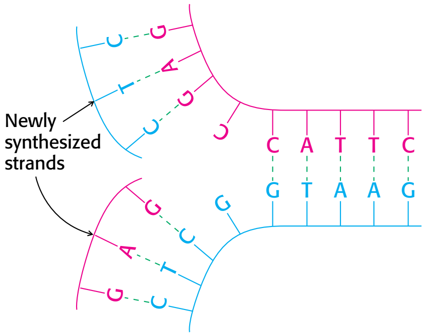


FIGURE 1.7 DNA replication. If a DNA molecule is separated into two strands, each strand can act as the template for the generation of its partner strand.

1.3 Concepts from Chemistry Explain the Properties of Biological Molecules

We have seen how a chemical insight into the hydrogen-bonding capabilities of the bases of DNA led to a deep understanding of a fundamental biological process. To lay the groundwork for the rest of the book, we begin our study of biochemistry by examining selected concepts from chemistry and showing how these concepts apply to biological systems. The concepts include the types of chemical bonds; the structure of water, the solvent in which most biochemical processes take place; the First and Second Laws of Thermodynamics; and the principles of acid–base chemistry.

The formation of the DNA double helix as a key example

We will use these concepts to examine an archetypical biochemical process—namely, the formation of a DNA double helix from its two component strands. The process is but one of many examples that could have been chosen to illustrate these topics. Keep in mind that, although the specific discussion is about DNA and double-helix formation, the concepts considered are quite general and will apply to many other classes of molecules and processes that will be discussed in the remainder of the book. In the course of these discussions, we will touch on the properties of water and the concepts of pK_a and buffers that are of great importance to many aspects of biochemistry.

The double helix can form from its component strands

The discovery that DNA from natural sources exists in a double-helical form with Watson–Crick base pairs suggested, but did not prove, that such double helices would form spontaneously outside biological systems. Suppose that two short strands of DNA were chemically synthesized to have complementary sequences so that they could, in principle, form a double helix with Watson–Crick base pairs. Two such sequences are CGATTAAT and ATTAATCG. The structures of these molecules in solution can be examined by a variety of techniques. In isolation, each sequence exists almost exclusively as a single-stranded molecule. However, when the two sequences are mixed, a double helix with Watson–Crick base pairs does form (Figure 1.8). This reaction proceeds nearly to completion.

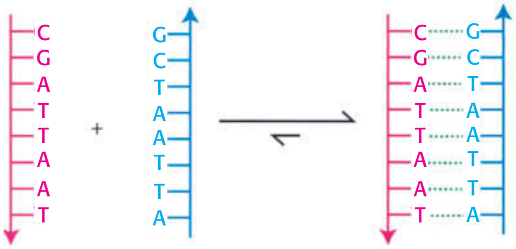


FIGURE 1.8 Formation of a double helix. When two DNA strands with appropriate, complementary sequences are mixed, they spontaneously assemble to form a double helix.

What forces cause the two strands of DNA to bind to each other? To analyze this binding reaction, we must consider several factors: the types of interactions and bonds in biochemical systems and the energetic favorability of the reaction. We must also consider the influence of the solution conditions—in particular, the consequences of acid–base reactions.

Covalent and noncovalent bonds are important for the structure and stability of biological molecules

Atoms interact with one another through chemical bonds. These bonds include the covalent bonds that define the structure of molecules as well as a variety of noncovalent bonds that are of great importance to biochemistry.

Covalent bonds. The strongest bonds are covalent bonds, such as the bonds that hold the atoms together within the individual bases shown on page 4. A covalent bond is formed by the sharing of a pair of electrons between adjacent atoms. A typical carbon–carbon (C—C) covalent bond has a bond length of 1.54 Å and bond energy of 355 kJ mol^{-1} (85 kcal mol^{-1}). Because covalent bonds are so strong, considerable energy must be expended to break them. More than one electron pair can be shared between two atoms to form a multiple covalent bond. For example, three of the bases in Figure 1.6 include carbon–oxygen (C═O) double bonds. These bonds are even stronger than C—C single bonds, with energies near 730 kJ mol^{-1} (175 kcal mol^{-1}) and are somewhat shorter.

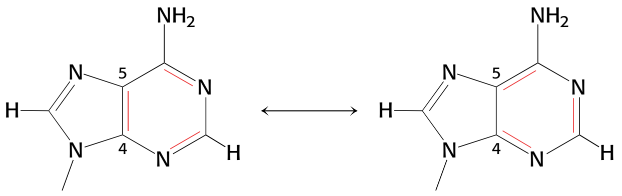
Distance and energy units

Interatomic distances and bond lengths are usually measured in angstrom (Å) units:

$$1\ \text{Å} = 10^{-10}\ \text{m} = 10^{-8}\ \text{cm} = 0.1\ \text{nm}$$

Several energy units are in common use. One joule (J) is the amount of energy required to move 1 meter against a force of 1 newton. A kilojoule (kJ) is 1000 joules. One calorie is the amount of energy required to raise the temperature of 1 gram of water 1 degree Celsius. A kilocalorie (kcal) is 1000 calories. One joule is equal to 0.239 cal.

For some molecules, more than one pattern of covalent bonding can be written. For example, adenine can be written in two nearly equivalent ways called *resonance structures.*

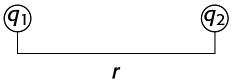


These adenine structures depict alternative arrangements of single and double bonds that are possible within the same structural framework. Resonance structures are shown connected by a double-headed arrow. Adenine's true structure is a composite of its two resonance structures. The composite structure is manifested in the bond lengths such as that for the bond joining carbon atoms C-4 and C-5. The observed bond length of 1.40 Å is between that expected for a C—C single bond (1.54 Å) and a C═C double bond (1.34 Å). A molecule that can be written as several resonance structures of approximately equal energies has greater stability than does a molecule without multiple resonance structures.

Noncovalent bonds. Noncovalent bonds are weaker than covalent bonds but are crucial for biochemical processes such as the formation of a double helix. Four fundamental noncovalent bond types are *ionic interactions, hydrogen bonds, van der Waals interactions,* and *hydrophobic interactions*. They differ in geometry, strength, and specificity. Furthermore, these bonds are affected in vastly different ways by the presence of water. Let us consider the characteristics of each type:

1. *Ionic Interactions*. A charged group on one molecule can attract an oppositely charged group on the same or another molecule. The energy of an ionic interaction (sometimes called an electrostatic interaction) is given by the *Coulomb energy:*

$$E = kq_1q_2/Dr$$



where E is the energy, q_1 and q_2 are the charges on the two atoms (in units of the electronic charge), r is the distance between the two atoms (in angstroms), D is the dielectric constant (which decreases the strength of the Coulomb depending on the intervening solvent or medium), and k is a proportionality constant ($k = 1389$, for energies in units of kilojoules per mole, or 332 for energies in kilocalories per mole).

By convention, an attractive interaction has a negative energy. The ionic interaction between two ions bearing single opposite charges separated by 3 Å in water (which has a dielectric constant of 80) has an energy of −5.8 kJ mol^{-1} (−1.4 kcal mol^{-1}). Note how important the dielectric constant of the medium is. For the same ions separated by 3 Å in a nonpolar solvent such as hexane (which has a dielectric constant of 2), the energy of this interaction is −232 kJ mol^{-1} (−55 kcal mol^{-1}).

2. *Hydrogen Bonds*. These interactions are largely ionic interactions, with partial charges on nearby atoms attracting one another. Hydrogen bonds are responsible for specific base-pair formation in the DNA double helix. The hydrogen atom in a hydrogen bond is partially shared by two electronegative atoms such as nitrogen or oxygen. The *hydrogen-bond donor* is the group that includes both the atom to which the hydrogen atom is more tightly linked and the hydrogen atom itself, whereas the *hydrogen-bond acceptor* is the atom less tightly linked to the hydrogen atom (Figure 1.9). The electronegative atom to which the hydrogen atom is covalently bonded pulls electron density away from the hydrogen atom, which thus develops a partial positive charge (δ^+). Thus, the hydrogen atom with a partial positive charge can interact with an atom having a partial negative charge (δ^-) through an ionic interaction.

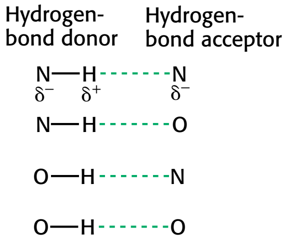
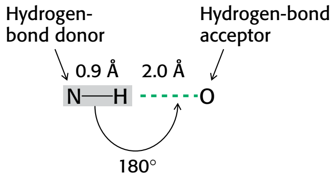


FIGURE 1.9 Hydrogen bonds. Hydrogen bonds are depicted by dashed green lines. The positions of the partial charges (δ^+ and δ^-) are shown.

Hydrogen bonds are much weaker than covalent bonds. They have energies ranging from 4 to 20 kJ mol^{-1} (from 1 to 5 kcal mol^{-1}). Hydrogen bonds are also somewhat longer than covalent bonds; their bond lengths (measured from the hydrogen atom) range from 1.5 Å to 2.6 Å; hence, a distance ranging from 2.4 Å to 3.5 Å separates the two nonhydrogen atoms in a hydrogen bond.



The strongest hydrogen bonds have a tendency to be approximately straight, such that the hydrogen-bond donor, the hydrogen atom, and the hydrogen-bond acceptor lie along a straight line. This tendency toward linearity can be important for orienting interacting molecules with respect to one another. Hydrogen-bonding interactions are responsible for many of the properties of water that make it such a special solvent, as will be described shortly.

3. *van der Waals Interactions*. The basis of a van der Waals interaction is that the distribution of electronic charge around an atom fluctuates with time. At any instant, the charge distribution is not perfectly symmetric. This transient asymmetry in the electronic charge about an atom acts through ionic interactions to induce a complementary asymmetry in the electron distribution within its neighboring atoms. The atom and its neighbors then attract one another. This attraction increases as two atoms come closer to each other, until they are separated by the van der Waals *contact distance* (Figure 1.10). At distances shorter than the van der Waals contact distance, very strong repulsive forces become dominant because the outer electron clouds of the two atoms overlap.

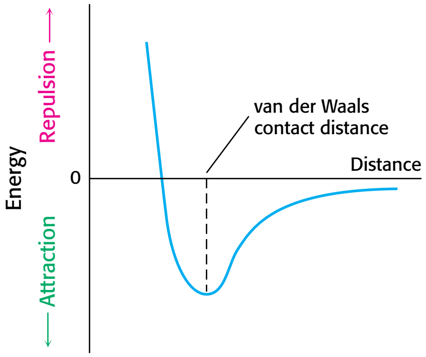
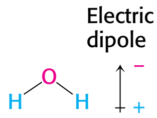


FIGURE 1.10 Energy of a van der Waals interaction as two atoms approach each other. The energy is most favorable at the van der Waals contact distance. Owing to electron–electron repulsion, the energy rises rapidly as the distance between the atoms becomes shorter than the contact distance.

Energies associated with van der Waals interactions are quite small; typical interactions contribute from 2 to 4 kJ mol^{-1} (from 0.5 to 1 kcal mol^{-1}) per atom pair. When the surfaces of two large molecules come together, however, a large number of atoms are in van der Waals contact, and the net effect, summed over many atom pairs, can be substantial.

We will cover the fourth noncovalent interaction, the hydrophobic interaction, after we examine the characteristics of water; these characteristics are essential to understanding the hydrophobic interaction.

Properties of water. Water is the solvent in which most biochemical reactions take place, and its properties are essential to the formation of macromolecular structures and the progress of chemical reactions. Two properties of water are especially relevant:



1. *Water is a polar molecule*. The water molecule is bent, not linear, and so the distribution of charge is asymmetric. The oxygen nucleus draws electrons away from the two hydrogen nuclei, which leaves the region around

each hydrogen atom with a net positive charge. The water molecule is thus an electrically polar structure.

2. *Water is highly cohesive*. Water molecules interact strongly with one another through hydrogen bonds. These interactions are apparent in the structure of ice (Figure 1.11). Networks of hydrogen bonds hold the structure together; similar interactions link molecules in liquid water and account for many of the properties of water. In the liquid state, approximately one in four of the hydrogen bonds present in ice are broken. The polar nature of water is responsible for its high dielectric constant of 80. Molecules in aqueous solution interact with water molecules through the formation of hydrogen bonds and through ionic interactions. These interactions make water a versatile solvent, able to readily dissolve many species, especially polar and charged compounds that can participate in these interactions.

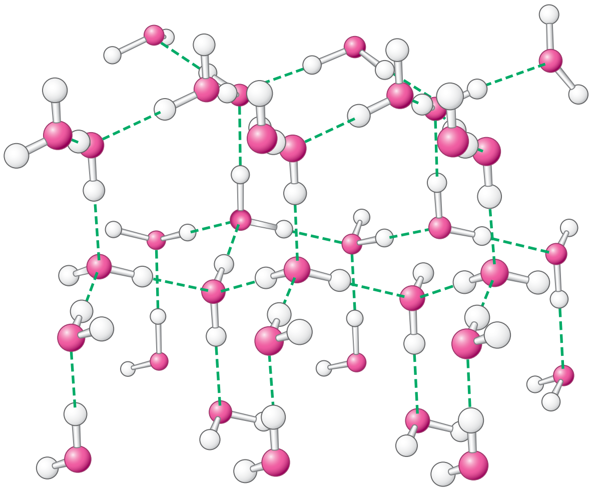


FIGURE 1.11 Structure of ice. Hydrogen bonds (shown as dashed green lines) are formed between water molecules to produce a highly ordered and open structure.

The hydrophobic effect. A final fundamental interaction called the *hydrophobic effect* is a manifestation of the properties of water. Some molecules (termed *nonpolar molecules*) cannot participate in hydrogen bonding or ionic interactions. The interactions of nonpolar molecules with water molecules are not as favorable as are interactions between the water molecules themselves. The water molecules in contact with these nonpolar molecules form "cages" around them, becoming more well ordered than water molecules free in solution. However, when two such nonpolar molecules come together, some of the water molecules are released, allowing them to interact freely with bulk water (Figure 1.12). The release of water from such cages is favorable for reasons to be considered shortly. The result is that nonpolar molecules show an increased tendency to associate with one another in water compared with other, less polar and less self-associating, solvents. This tendency is called the hydrophobic effect and the associated interactions are called *hydrophobic interactions*.

The double helix is an expression of the rules of chemistry

Let us now see how these four noncovalent interactions work together in driving the association of two strands of DNA to form a double helix. First, each phosphate group in a DNA strand carries a negative charge. These negatively charged groups interact unfavorably with one another over distances. Thus, unfavorable ionic interactions take place when two strands of

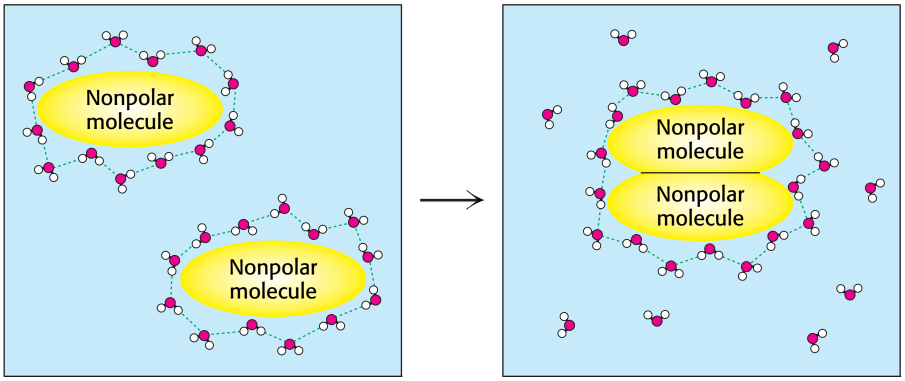


FIGURE 1.12 The hydrophobic effect. The aggregation of nonpolar groups in water leads to the release of water molecules, initially interacting with the nonpolar surface, into bulk water. The release of water molecules into solution makes the aggregation of nonpolar groups favorable.

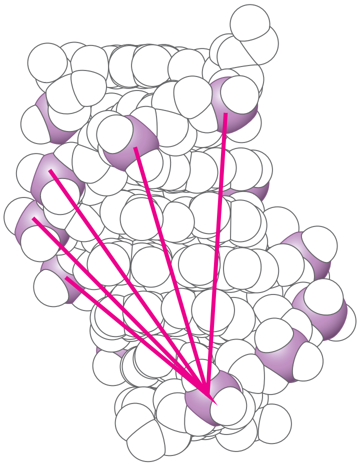
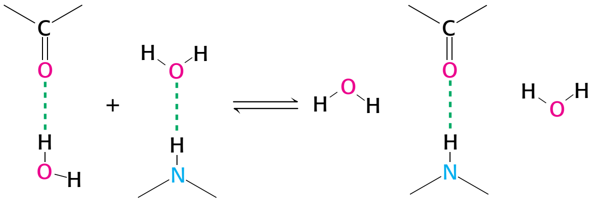


FIGURE 1.13 Ionic interactions in DNA. Each unit within the double helix includes a phosphate group (the phosphorus atom being shown in purple) that bears a negative charge. The unfavorable interactions of one phosphate with several others are shown by red lines. These repulsive interactions oppose the formation of a double helix.

DNA come together. These phosphate groups are far apart in the double helix with distances greater than 10 Å, but many such interactions take place (Figure 1.13). Thus, ionic interactions oppose the formation of the double helix. The strength of these repulsive ionic interactions is diminished by the high dielectric constant of water and the presence of ionic species such as Na^+ or Mg^{2+} ions in solution. These positively charged species interact with the phosphate groups and partly neutralize their negative charges.

Second, as already noted, hydrogen bonds are important in determining the formation of specific base pairs in the double helix. However, in single-stranded DNA, the hydrogen-bond donors and acceptors are exposed to solution and can form hydrogen bonds with water molecules.



When two single strands come together, these hydrogen bonds with water are broken and new hydrogen bonds between the bases are formed. Because the number of hydrogen bonds broken is the same as the number formed, these hydrogen bonds do not contribute substantially to driving the overall process of double-helix formation. However, they contribute greatly to the specificity of binding. Suppose two bases that cannot form Watson–Crick base pairs are brought together. Hydrogen bonds with water must be broken as the bases come into contact. Because the bases are not complementary in structure, not all of these bonds can be simultaneously replaced by hydrogen bonds between the bases. Thus, the formation of a double helix between noncomplementary sequences is disfavored.

Third, within a double helix, the base pairs are parallel and stacked nearly on top of one another. The typical separation between the planes of adjacent base pairs is 3.4 Å, and the distances between the most closely approaching atoms are approximately 3.6 Å. This separation distance corresponds nicely to the van der Waals contact distance (Figure 1.14). Bases tend to stack even in single-stranded DNA molecules. However, the base stacking and associated van der Waals interactions are nearly optimal in a double-helical structure.

Fourth, the hydrophobic effect also contributes to the favorability of base stacking. More-complete base stacking moves the nonpolar surfaces of the bases out of water into contact with each other.

The principles of double-helix formation between two strands of DNA apply to many other biochemical processes. Many weak interactions contribute to the overall energetics of the process, some favorably and some unfavorably. Furthermore, surface complementarity is a key feature: when complementary surfaces meet, hydrogen-bond donors align with hydrogen-bond acceptors and nonpolar surfaces come together to maximize van der Waals interactions and minimize nonpolar surface area exposed to the aqueous environment. The properties of water play a major role in determining the importance of these interactions.

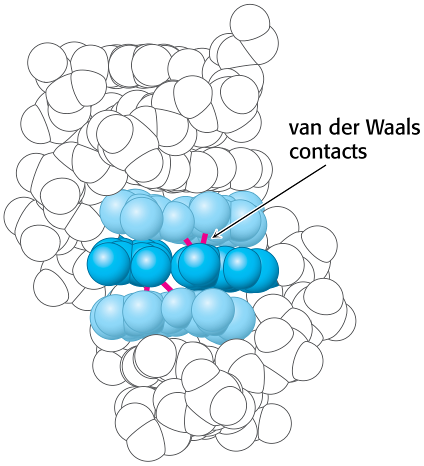


FIGURE 1.14 Base stacking. In the DNA double helix, adjacent base pairs are stacked nearly on top of one another, and so many atoms in each base pair are separated by their van der Waals contact distance. The central base pair is shown in dark blue and the two adjacent base pairs in light blue. Several van der Waals contacts are shown in red.

The laws of thermodynamics govern the behavior of biochemical systems

We can look at the formation of the double helix from a different perspective by examining the laws of thermodynamics. These laws are general principles that apply to all physical (and biological) processes. They are of great importance because they determine the conditions under which specific processes can or cannot take place. We will consider these laws from a general perspective first and then apply the principles that we have developed to the formation of the double helix.

The laws of thermodynamics distinguish between a system and its surroundings. A *system* refers to the matter within a defined region of space. The matter in the rest of the universe is called the *surroundings. The First Law of Thermodynamics states that the total energy of a system and its surroundings is constant*. In other words, the energy content of the universe is constant; energy can be neither created nor destroyed. Energy can take different forms, however. Heat, for example, is one form of energy. Heat is a manifestation of the *kinetic energy* associated with the random motion of molecules. Alternatively, energy can be present as *potential energy*—energy that will be released on the occurrence of some process. Consider, for example, a ball held at the top of a tower. The ball has considerable potential energy because, when it is released, the ball will develop kinetic energy associated with its motion as it falls. Within chemical systems, potential energy is related to the likelihood that atoms can react with one another. For instance, a mixture of gasoline and oxygen has a large potential energy because these molecules may react to form carbon dioxide and water and release energy as heat. The First Law requires that any energy released in the formation of chemical bonds must be used to break other bonds, released as heat or light, or stored in some other form.

Another important thermodynamic concept is that of *entropy*, a measure of the degree of randomness or disorder in a system. *The Second Law of Thermodynamics states that the total entropy of a system plus that of its surroundings always increases*. For example, the release of water from nonpolar surfaces responsible for the hydrophobic effect is favorable because water molecules free in solution are more disordered than they are when they are associated with nonpolar surfaces. At first glance, the Second Law appears to contradict much common experience, particularly about biological systems. Many biological processes, such as the generation of a leaf from carbon dioxide gas and other nutrients, clearly increase the level of order and hence decrease entropy. Entropy may be decreased locally in the formation of such ordered structures only if the entropy of other parts of the universe is increased by an equal or greater amount. The local decrease in entropy is often accomplished by a release of heat, which increases the entropy of the surroundings.

We can analyze this process in quantitative terms. First, consider the system. The entropy (S) of the system may change in the course of a chemical reaction by an amount ΔS_{system}. If heat flows from the system to its surroundings, then the heat content, often referred to as the *enthalpy* (H), of the system will be reduced by an amount ΔH_{system}. To apply the Second Law, we must determine the change in entropy of the surroundings. If heat flows from the system to the surroundings, then the entropy of the surroundings will increase. The precise change in the entropy of the surroundings depends on the temperature; the change in entropy is greater when heat is added to relatively cold surroundings than when heat is added to surroundings at high temperatures that are already in a high degree of disorder. To be even more specific, the change in the entropy of the surroundings will be proportional to the amount of heat transferred from the system and inversely proportional to the temperature (T) of the surroundings. In biological systems, T [in kelvins (K), absolute temperature] is usually assumed to be constant. Thus, a change in the entropy of the surroundings is given by

$$\Delta S_{\text{surroundings}} = -\Delta H_{\text{system}}/T \tag{1}$$

The total entropy change is given by the expression

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{surroundings}} \tag{2}$$

Substituting equation 1 into equation 2 yields

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} - \Delta H_{\text{system}}/T \tag{3}$$

Multiplying by $-T$ gives

$$-T\Delta S_{\text{total}} = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \tag{4}$$

The function $-T\Delta S$ has units of energy and is referred to as *free energy* or *Gibbs free energy,* after Josiah Willard Gibbs, who developed this function in 1878:

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \tag{5}$$

The free-energy change, ΔG, will be used throughout this book to describe the energetics of biochemical reactions. The Gibbs free energy is essentially an accounting tool that keeps track of both the entropy of the system (directly) and the entropy of the surroundings (in the form of heat released from the system).

Recall that the Second Law of Thermodynamics states that, for a process to take place, the entropy of the universe must increase. Examination of equation 3 shows that the total entropy will increase if and only if

$$\Delta S_{\text{system}} > \Delta H_{\text{system}}/T \tag{6}$$

Rearranging gives $T\Delta S_{\text{system}}$. ΔH or, in other words, entropy will increase if and only if

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} < 0 \tag{7}$$

Thus, the free-energy change must be negative for a process to take place spontaneously. *There is negative free-energy change when and only when the overall entropy of the universe is increased*. Again, the free energy represents a single term that takes into account both the entropy of the system and the entropy of the surroundings.

Heat is released in the formation of the double helix

Let us see how the principles of thermodynamics apply to the formation of the double helix (Figure 1.15). Suppose solutions containing each of the two single strands are mixed. Before the double helix forms, each of the single strands is free to translate and rotate in solution, whereas each matched pair of strands in the double helix must move together. Furthermore, the free single strands exist in more conformations than possible when bound together in a double helix. Thus, the formation of a double helix from two single strands appears to result in an increase in order for the system, that is, a decrease in the entropy of the system.

On the basis of this analysis, we expect that the double helix cannot form without violating the Second Law of Thermodynamics unless heat is released to increase the entropy of the surroundings. Experimentally, we can measure the heat released by allowing the solutions containing the two single strands to come together within a water bath, which here corresponds to the surroundings. We then determine how much heat must be absorbed by the water bath or released from it to maintain it at a constant temperature. This experiment reveals that a substantial amount of heat is released—namely, approximately 250 kJ mol^{-1} (60 kcal mol^{-1}). This experimental result reveals that the change in enthalpy for the process is quite large, -250 kJ mol^{-1}, consistent with our expectation that significant heat would have to be released to the surroundings for the process not to violate the Second Law. We see in quantitative terms how order within a system can be increased by releasing sufficient heat to the surroundings to ensure that the entropy of the universe increases. We will encounter this general theme again and again throughout this book.

Acid–base reactions are central in many biochemical processes

Throughout our consideration of the formation of the double helix, we have dealt only with the noncovalent bonds that are formed or broken in this process. Many biochemical processes entail the formation and cleavage of covalent bonds. A particularly important class of reactions prominent in biochemistry is *acid–base reactions*.

In acid and base reactions, hydrogen ions are added to molecules or removed from them. Throughout the book, we will encounter many processes in which the addition or removal of hydrogen atoms is crucial, such as the metabolic processes by which carbohydrates are degraded to release energy for other uses. Thus, a thorough understanding of the basic principles of these reactions is essential.

A hydrogen ion, often written as H^+, corresponds to a proton. In fact, hydrogen ions exist in solution bound to water molecules, thus forming what are known as *hydronium ions,* H_3O^+. For simplicity, we will continue to write H^+, but we should keep in mind that H^+ is shorthand for the actual species present.

The concentration of hydrogen ions in solution is expressed as the pH. Specifically, the *pH* of a solution is defined as

$$\mathrm{pH} = -\log[\mathrm{H}^+]$$

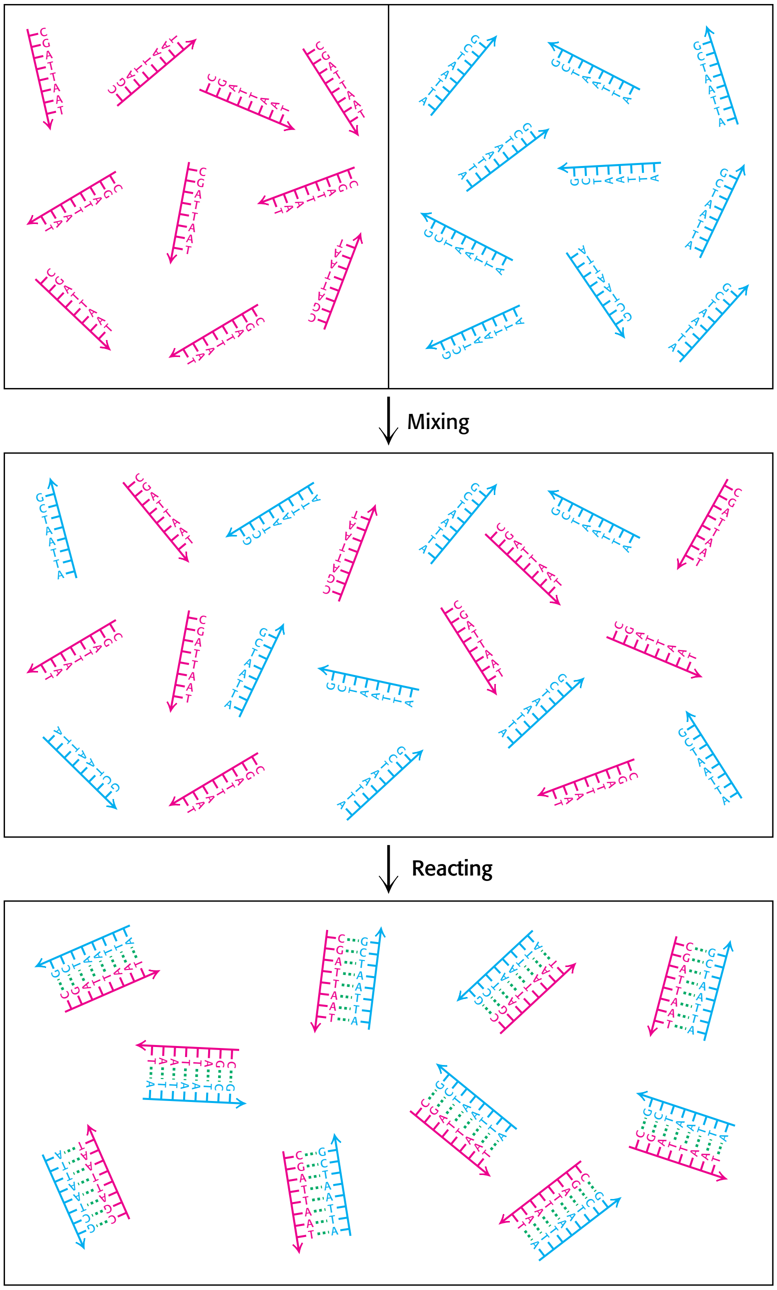


FIGURE 1.15 Double-helix formation and entropy. When solutions containing DNA strands with complementary sequences are mixed, the strands react to form double helices. This process results in a loss of entropy from the system, indicating that heat must be released to the surroundings to prevent a violation of the Second Law of Thermodynamics.

where $[\mathrm{H}^+]$ is in units of molarity. Thus, pH 7.0 refers to a solution for which $-\log[\mathrm{H}^+] = 7.0$, and so $\log[\mathrm{H}^+] = -7.0$ and $[\mathrm{H}^+] = 10^{\log[\mathrm{H}+]} = 10^{-7.0} = 1.0 \times 10^{-7}$ M.

The pH also indirectly expresses the concentration of hydroxide ions, $[\mathrm{OH}^-]$, in solution. To see how, we must realize that water molecules can dissociate to form H^+ and OH^- ions in an equilibrium process.

$$\mathrm{H_2O} \rightleftharpoons \mathrm{H}^+ + \mathrm{OH}^-$$

The equilibrium constant (K) for the dissociation of water is defined as

$$K = [\mathrm{H}^+][\mathrm{OH}^-]/[\mathrm{H_2O}]$$

and has a value of $K = 1.8 \times 10^{-16}$. Note that an equilibrium constant does not formally have units. Nonetheless, the value of the equilibrium constant given assumes that particular units are used for concentration (sometimes referred to a standard states); in this case and in many others, units of molarity (M) are assumed.

The concentration of water, $[H_2O]$, in pure water is 55.5 M, and this concentration is constant under most conditions. Thus, we can define a new constant, K_W:

$$K_W = K[H_2O] = [H^+][OH^-]$$

$$K[H_2O] = 1.8 \times 10^{-16} \times 55.5$$

$$= 1.0 \times 10^{-14}$$

Because $K_W = [H^+][OH^-] = 1.0 \times 10^{-14}$, we can calculate

$$[OH^-] = 10^{-14}/[H^+] \text{ and } [H^+] = 10^{-14}/[OH^-]$$

With these relations in hand, we can easily calculate the concentration of hydroxide ions in an aqueous solution, given the pH. For example, at pH = 7.0, we know that $[H^+] = 10^{-7}$ M and so $[OH^-] = 10^{-14}/10^{-7} = 10^{-7}$ M. In acidic solutions, the concentration of hydrogen ions is higher than 10^{-7} and, hence, the pH is below 7. For example, in 0.1 M HCl, $[H^+] = 10^{-1}$ M and so pH = 1.0 and $[OH^-] = 10^{-14}/10^{-1} = 10^{-13}$ M.

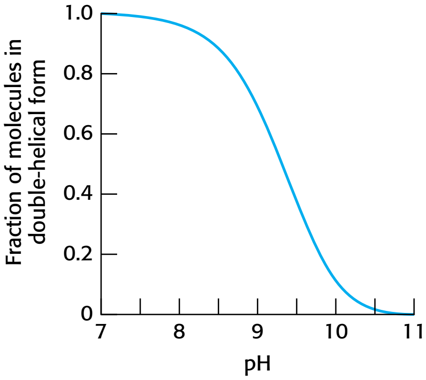
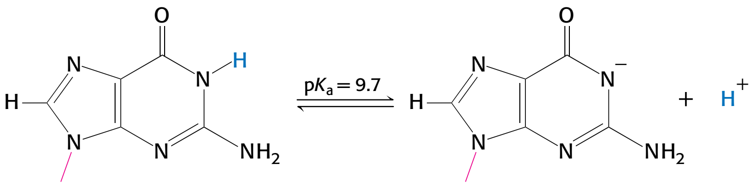


FIGURE 1.16 DNA denaturation by the addition of a base. The addition of a base to a solution of double-helical DNA initially at pH 7 causes the double helix to separate into single strands. The process is half complete at slightly above pH 9.

Acid–base reactions can disrupt the double helix

The reaction that we have been considering between two strands of DNA to form a double helix takes place readily at pH 7.0. Suppose that we take the solution containing the double-helical DNA and treat it with a solution of concentrated base (i.e., with a high concentration of OH^-). As the base is added, we monitor the pH and the fraction of DNA in double-helical form (Figure 1.16). When the first additions of base are made, the pH rises, but the concentration of the double-helical DNA does not change significantly. However, as the pH approaches 9, the DNA double helix begins to dissociate into its component single strands. As the pH continues to rise from 9 to 10, this dissociation becomes essentially complete. Why do the two strands dissociate? The hydroxide ions can react with bases in DNA base pairs to remove certain protons. The most susceptible proton is the one bound to the N-1 nitrogen atom in a guanine base.



Guanine (G)

Proton dissociation for a substance HA (such as that bound to N-1 on guanine) has an equilibrium constant defined by the expression

$$K_a = [H^+][A^-]/[HA]$$

The susceptibility of a proton to removal by reaction with a base is often described by its *pK_a value:*

$$pK_a = -\log(K_a)$$

When the pH is equal to the pK_a, we have

$$\text{pH} = \text{p}K_a$$

and so

$$-\log[\text{H}^+] = -\log([\text{H}^+][\text{A}^-]/[\text{HA}])$$

and

$$[\text{H}^+] = [\text{H}^+][\text{A}^-]/[\text{HA}]$$

Dividing by $[\text{H}^+]$ reveals that

$$1 = [\text{A}^-]/[\text{HA}]$$

and so

$$[\text{A}^-] = [\text{HA}]$$

Thus, when the pH equals the pK_a, the concentration of the deprotonated form of the group or molecule is equal to the concentration of the protonated form; the deprotonation process is halfway to completion.

The pK_a for the proton on N-1 of guanine is typically 9.7. When the pH approaches this value, the proton on N-1 is lost (Figure 1.16). Because this proton participates in an important hydrogen bond, its loss substantially destabilizes the DNA double helix. The DNA double helix is also destabilized by *low* pH. Below pH 5, some of the hydrogen bond *acceptors* that participate in base-pairing become protonated. In their protonated forms, these bases can no longer form hydrogen bonds and the double helix separates. Thus, acid–base reactions that remove or donate protons at specific positions on the DNA bases can disrupt the double helix.

Buffers regulate pH in organisms and in the laboratory

These observations about DNA reveal that a significant change in pH can disrupt molecular structure. The same is true for many other biological macromolecules; changes in pH can protonate or deprotonate key groups, potentially disrupting structures and initiating harmful reactions. Thus, systems have evolved to mitigate changes in pH in biological systems. Solutions that resist such changes are called *buffers*. Specifically, when acid is added to an unbuffered aqueous solution, the pH drops in proportion to the amount of acid added. In contrast, when acid is added to a buffered solution, the pH drops more gradually. Buffers also mitigate the pH increase caused by the addition of base and changes in pH caused by dilution.

Compare the result of adding a 1 M solution of the strong acid HCl drop by drop to pure water with adding it to a solution containing 100 mM of the buffer sodium acetate ($\text{Na}^+\text{CH}_3\text{COO}^-$; Figure 1.17). The process of gradually adding known amounts of reagent to a solution with which the reagent reacts while monitoring the results is called a *titration*. For pure water, the pH drops from 7 to close to 2 on the addition of the first few drops of acid. However, for the sodium acetate solution, the pH first falls rapidly from its initial value near 10, then changes more gradually until the pH reaches 3.5, and then falls more rapidly again. Why does the pH decrease so gradually in the middle of the titration? The answer is that, when hydrogen ions are added to this solution, they react with acetate ions to form acetic acid. This reaction consumes some of the added hydrogen ions so that the pH does not drop. Hydrogen ions continue reacting with acetate ions until essentially all of the acetate ion is converted into acetic acid. After this point, added protons remain free in solution and the pH begins to fall sharply again.

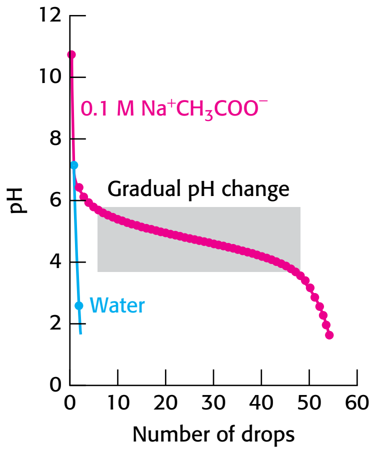


FIGURE 1.17 Buffer action. The addition of a strong acid, 1 M HCl, to pure water results in an immediate drop in pH to near 2. In contrast, the addition of the acid to a 0.1 M sodium acetate (Na^+ CH_3COO^-) solution results in a much more gradual change in pH until the pH drops below 3.5.

We can analyze the effect of the buffer in quantitative terms. The equilibrium constant for the deprotonation of an acid is

$$K_a = [H^+][A^-]/[HA]$$

Taking logarithms of both sides yields

$$\log(K_a) = \log([H^+]) + \log([A^-]/[HA])$$

Recalling the definitions of pK_a and pH and rearranging gives

$$pH = pK_a + \log([A^-]/[HA])$$

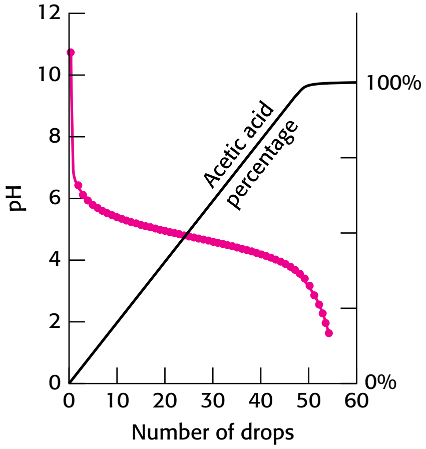


FIGURE 1.18 Buffer protonation. When acid is added to sodium acetate, the added hydrogen ions are used to convert acetate ion into acetic acid. Because the proton concentration does not increase significantly, the pH remains relatively constant until all of the acetate has been converted into acetic acid.

This expression is referred to as the *Henderson–Hasselbalch equation.*

We can apply the equation to our titration of sodium acetate. The pK_a of acetic acid is 4.75. We can calculate the ratio of the concentration of acetate ion to the concentration of acetic acid as a function of pH by using the Henderson–Hasselbalch equation, slightly rearranged.

$$[\text{Acetate ion}]/[\text{Acetic acid}] = [A^-]/[HA] = 10^{pH-pK_a}$$

At pH 9, this ratio is $10^{9-4.75} = 10^{4.25} = 17{,}800$; very little acetic acid has been formed. At pH 4.75 (when the pH equals the pK_a), the ratio is $10^{4.75-4.75} = 10^0 = 1$. At pH 3, the ratio is $10^{3-4.75} = 10^{-1.25} = 0.02$; almost all of the acetate ion has been converted into acetic acid. We can follow the conversion of acetate ion into acetic acid over the entire titration (Figure 1.18). The graph shows that the region of relatively constant pH corresponds precisely to the region in which acetate ion is being protonated to form acetic acid.

From this discussion, we see that a buffer functions best close to the pK_a value of its acid component. Physiological pH is typically about 7.4. An important buffer in biological systems is based on phosphoric acid (H_3PO_4). The acid can be deprotonated in three steps to form a phosphate ion.

$$H_3PO_4 \xrightleftharpoons[pK_a = 2.12]{H^+} H_2PO_4^- \xrightleftharpoons[pK_a = 7.21]{H^+} HPO_4^{2-} \xrightleftharpoons[pK_a = 12.67]{H^+} PO_4^{3-}$$

At about pH 7.4, inorganic phosphate exists primarily as a nearly equal mixture of $H_2PO_4^-$ and HPO_4^{2-}. Thus, phosphate solutions function as effective buffers near pH 7.4. The concentration of inorganic phosphate in blood is typically approximately 1 mM, providing a useful buffer against processes that produce either acid or base. We can examine this utility in quantitative terms with the use of the Henderson–Hasselbalch equation. What concentration of acid must be added to change the pH of 1 mM phosphate buffer from 7.4 to 7.3? Without buffer, this change in $[H^+]$ corresponds to a change of $10^{-7.3} - 10^{-7.4}$ M $= (5.0 \times 10^{-8} - 4.0 \times 10^{-8})$ M $= 1.0 \times 10^{-8}$ M. Let us now consider what happens to the buffer components. At pH 7.4,

$$[HPO_4^{2-}]/[H_2PO_4^-] = 10^{7.4-7.21} = 10^{0.19} = 1.55$$

The total concentration of phosphate, $[HPO_4^{2-}] + [H_2PO_4^-]$, is 1 mM, Thus,

$$[HPO_4^{2-}] = (1.55/2.55) \times 1\text{ mM} = 0.608\text{ mM}$$

and

$$[H_2PO_4^-] = (1/2.55) \times 1\text{ mM} = 0.392\text{ mM}$$

At pH 7.3,

$$[HPO_4{}^{2-}]/[H_2PO_4{}^{-}] = 10^{7.3-7.21} = 10^{0.09} = 1.23$$

and so

$$[HPO_4{}^{2-}] = (1.23/2.23) = 0.552\ \text{mM}$$

and

$$[H_2PO_4{}^{-}] = (1/2.23) = 0.448\ \text{mM}$$

Thus, $(0.608 - 0.552) = 0.056$ mM $HPO_4{}^{2-}$ is converted into $H_2PO_4{}^{-}$, consuming 0.056 mM $= 5.6 \times 10^{-5}$ M $[H^+]$. Thus, the buffer increases the amount of acid required to produce a drop in pH from 7.4 to 7.3 by a factor of $5.6 \times 10^{-5}/1.0 \times 10^{-8} = 5600$ compared with pure water.

1.4 The Genomic Revolution Is Transforming Biochemistry, Medicine, and Other Fields

Watson and Crick's discovery of the structure of DNA suggested the hypothesis that hereditary information is stored as a sequence of bases along long strands of DNA. This remarkable insight provided an entirely new way of thinking about biology. However, at the time that it was made, Watson and Crick's discovery, though full of potential, remained to be confirmed and many features needed to be elucidated. How is the sequence information read and translated into action? What are the sequences of naturally occurring DNA molecules and how can such sequences be experimentally determined? Through advances in biochemistry and related sciences, we now have essentially complete answers to these questions. Indeed, in the past decade or so, scientists have determined the complete genome sequences of hundreds of different organisms, including simple microorganisms, plants, animals of varying degrees of complexity, and human beings. Comparisons of these genome sequences, with the use of methods introduced in Chapter 6, have been sources of many insights that have transformed biochemistry. In addition to its experimental and clinical aspects, biochemistry has now become an *information science*.

Genome sequencing has transformed biochemistry and other fields

The sequencing of a human genome was a daunting task because it contains approximately 3 billion (3×10^9) base pairs. For example, the sequence

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTC
AAACAGACACCATGGTGCATCTGACTCCTG**A**GGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGA . . .

is a part of one of the genes that encodes hemoglobin, the oxygen carrier in our blood. This gene is found on the end of chromosome 9 of our 24 distinct chromosomes. If we were to include the complete sequence of our entire genome, this chapter would run to more than 500,000 pages. The sequencing of our genome is truly a landmark in human history. This sequence contains a vast amount of information, some of which we can now extract and interpret, but much of which we are only beginning to understand. For example, some human diseases have been linked to particular variations in genomic sequence. Sickle-cell anemia, discussed in detail in Chapter 7, is caused by a single base change of an A (noted in boldface type in the preceding sequence) to a T. We will encounter many other examples of diseases that have been linked to specific DNA sequence changes.

Determining the first human genome sequences was a great challenge. It required the efforts of large teams of geneticists, molecular biologists, biochemists, and computer scientists, as well as billions of dollars, because there was no previous framework for aligning the sequences of various DNA fragments. One human genome sequence can serve as a reference for other sequences. The availability of such reference sequences enables much more rapid characterization of partial or complete genomes from other individuals. As we will discuss in Chapter 5, arrays containing millions of target single-stranded DNA molecules with sequences from the reference genome and known or potential variants are powerful tools. These arrays can be exposed to mixtures of DNA fragments for a particular individual and those single-stranded targets that bind to their complementary strands can be determined. This allows many positions within the genome of the individual to be simulaneously interrogated.

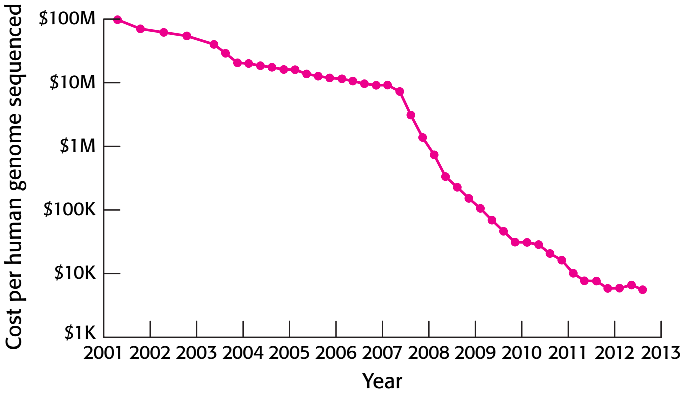


FIGURE 1.19 Decreasing costs of DNA sequencing. Through the Human Genome Project, the cost of DNA sequencing dropped steadily. With the advent of new methods, these costs dropped dramatically and are now approaching $1000 for a complete human genome sequence. [National Human Genome Research Institute. www.genome.gov/sequencingcosts]

Methods for sequencing DNA have also been improving rapidly, driven by a deep understanding of the biochemistry of DNA replication and other processes. This has resulted in both dramatic increases in the DNA sequencing rate and decreases in cost (Figure 1.19). The availability of such powerful sequencing technology is transforming many fields, including medicine, dentistry, microbiology, pharmacology, and ecology, although a great deal remains to be done to improve the accuracy and precision of the interpretation of these large genomic and related data sets.

Each person has a unique sequence of DNA base pairs. How different are we from one another at the genomic level? Examination of genomic variation reveals that, on average, each pair of individuals has a different base in one position per 200 bases; that is, the difference is approximately 0.5%. This variation between individuals who are not closely related is quite substantial compared with differences in populations. The average difference between two people within one ethnic group is greater than the difference between the averages of two different ethnic groups.

The significance of much of this genetic variation is not understood. As noted earlier, variation in a single base within the genome can lead to a disease such as sickle-cell anemia. Scientists have now identified the genetic variations associated with hundreds of diseases for which the cause can be traced to a single gene. For other diseases and traits, we know that variation in many different genes contributes in significant and often complex ways. Many of the most prevalent human ailments such as heart disease are linked to variations in many genes. Furthermore, in most cases, the presence of a particular variation or set of variations does not inevitably result in the onset of a disease but, instead, leads to a *predisposition* to the development of the disease.

Our own genes are not the only ones that can contribute to health and disease. Our bodies, including our skin, mouth, digestive tract, genitourinary tract, respiratory tract, and other areas, contain large number of microorganisms. These complex communities have been characterized through powerful methods that allow DNA isolated from these biological samples to be sequenced without any previous knowledge of the organisms present. Many of these organisms had not previously been discovered because they can only grow as part of complex communities and thus cannot be isolated through standard microbiological techniques. Remarkably, it appears that we are outnumbered in our own bodies! Each of us contains approximately ten times more microbial cells than human cells and these microbial cells include many more genes than do our own genomes. These *microbiomes* differ from site to site, from one person to another and over time in the same individual. They appear to play roles in health and in diseases such as obesity and dental caries (Figure 1.20).

In addition to the implications for understanding human health and disease, the genome sequence is a source of deep insight into other aspects of human biology and culture. For example, by comparing the sequences of different individuals and populations, we can learn a great deal about human history. On the basis of such analysis, a compelling case can be made that the human species originated in Africa, and the occurrence and even the timing of important migrations of groups of human beings can be demonstrated (Figure 1.21). Finally, comparisons of the human genome with the genomes of other organisms are confirming the tremendous unity that exists at the level of biochemistry and are revealing key steps in the course of evolution from relatively simple, single-celled organisms to complex, multicellular organisms such as human beings. For example, many genes that are key to the function of the human brain and nervous system have evolutionary and functional relatives in the genomes of bacteria. Because many studies that are possible in model organisms are difficult or unethical to conduct in human beings, these discoveries have many practical implications. *Comparative genomics* has become a powerful science, linking evolution and biochemistry.

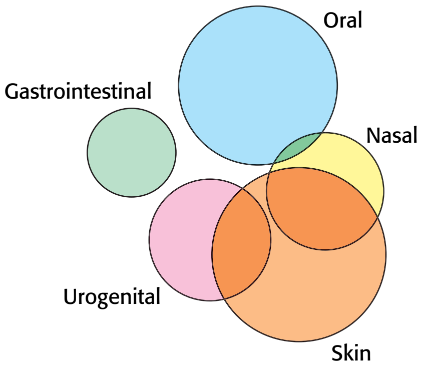


FIGURE 1.20 The human microbiome. Microorganisms cover the human body. Examination of the microbial communities using DNA sequencing methods revealed many previously uncharacterized species. The Venn diagrams represent populations of related species as determined by DNA sequence comparisons. The populations present on different body surfaces are largely distinct. [Adapted from www.nature.com/nature/journal/v486/n7402/fig_tab/nature11234_F1.html]

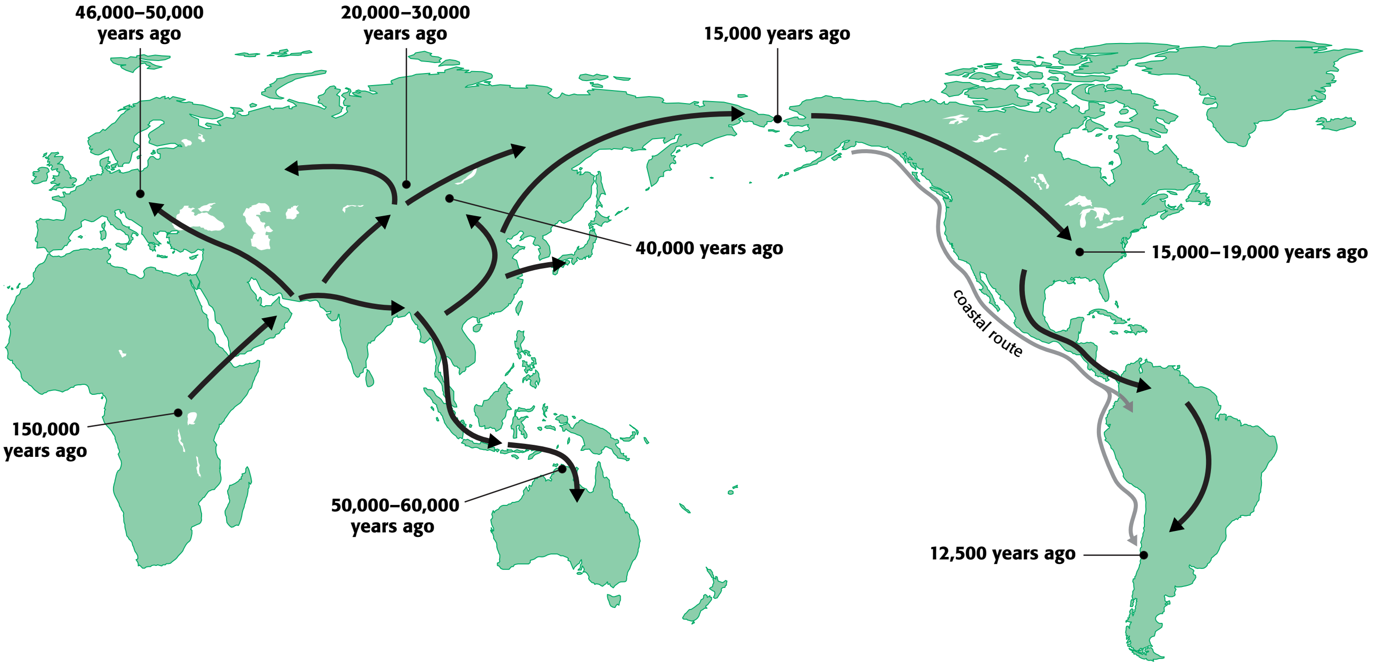


FIGURE 1.21 Human migrations supported by DNA sequence comparisons. Modern human beings originated in Africa, migrated first to Asia, and then to Europe, Australia, and North and South America. [Adapted from S. Oppenheimer, "Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map." *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367(1590):770–784]

Environmental factors influence human biochemistry

Although our genetic makeup (and that of our microbiomes) is an important factor that contributes to disease susceptibility and to other traits, factors in a person's environment also are significant. What are these environmental factors? Perhaps the most obvious are chemicals that we eat or are exposed to in some other way. The adage "you are what you eat" has considerable validity; it applies both to substances that we ingest in significant quantities and to those that we ingest in only trace amounts. Throughout our study of biochemistry, we will encounter *vitamins* and *trace elements* and their derivatives that play crucial roles in many processes. In many cases, the roles of these chemicals were first revealed through investigation of *deficiency disorders* observed in people who do not take in a sufficient quantity of a particular vitamin or trace element. Despite the fact that the most important essential dietary factors have been known for some time, new roles for them continue to be discovered.

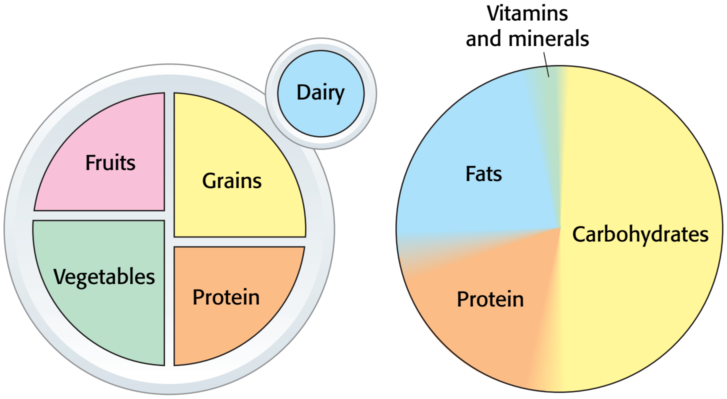


FIGURE 1.22 Nutrition. Proper health depends on an appropriate combination of food groups (fruits, vegetables, protein, grains, dairy) (left) to provide an optimal mix of biochemicals (carbohydrates, proteins, fats, vitamins, and minerals) (right). [Adapted from www.choosemyplate.gov]

A healthful diet requires a balance of major food groups. In addition to providing vitamins and trace elements, food provides calories in the form of substances that can be broken down to release energy that drives other biochemical processes. Proteins, fats, and carbohydrates provide the building blocks used to construct the molecules of life (Figure 1.22). Finally, it is possible to get too much of a good thing. Human beings evolved under circumstances in which food, particularly rich foods such as meat, was scarce. With the development of agriculture and modern economies, rich foods are now plentiful in parts of the world. Some of the most prevalent diseases in the so-called developed world, such as heart disease and diabetes, can be attributed to the large quantities of fats and carbohydrates present in modern diets. We are now developing a deeper understanding of the biochemical consequences of these diets and the interplay between diet and genetic factors.

Chemicals are only one important class of environmental factors. Our behaviors also have biochemical consequences. Through physical activity, we consume the calories that we take in, ensuring an appropriate balance between food intake and energy expenditure. Activities ranging from exercise to emotional responses such as fear and love may activate specific biochemical pathways, leading to changes in levels of gene expression, the release of hormones, and other consequences. Furthermore, the interplay between biochemistry and behavior is bidirectional. Just as our biochemistry is affected by our behavior, so, too, our behavior is affected, although certainly not completely determined, by our genetic makeup and other aspects of our biochemistry. Genetic factors associated with a range of behavioral characteristics have been at least tentatively identified.

Just as vitamin deficiencies and genetic diseases have revealed fundamental principles of biochemistry and biology, investigations of variations in behavior and their linkage to genetic and biochemical factors are potential sources of great insight into mechanisms within the brain. For example, studies of drug addiction have revealed neural circuits and biochemical pathways that greatly influence aspects of behavior. Unraveling the interplay between biology and behavior is one of the great challenges in modern science, and biochemistry is providing some of the most important concepts and tools for this endeavor.

Genome sequences encode proteins and patterns of expression

The structure of DNA revealed how information is stored in the base sequence along a DNA strand. But what information is stored and how is it expressed? The most fundamental role of DNA is to encode the sequences of proteins. Like DNA, proteins are linear polymers. However, proteins differ from DNA in two important ways. First, proteins are built from 20 building blocks, called *amino acids,* rather than just four, as in DNA. The chemical complexity provided by this variety of building blocks enables proteins to perform a wide range of functions. Second, proteins spontaneously fold into elaborate three-dimensional structures, determined only by their amino acid sequences (Figure 1.23). We have explored in depth how solutions containing two appropriate strands of DNA come together to form a solution of double-helical molecules. A similar spontaneous folding process gives proteins their three-dimensional structure. A balance of hydrogen bonding, van der Waals interactions, and hydrophobic interactions overcomes the entropy lost in going from an unfolded ensemble of proteins to a homogenous set of well-folded molecules. Proteins and protein folding will be discussed extensively in Chapter 2.

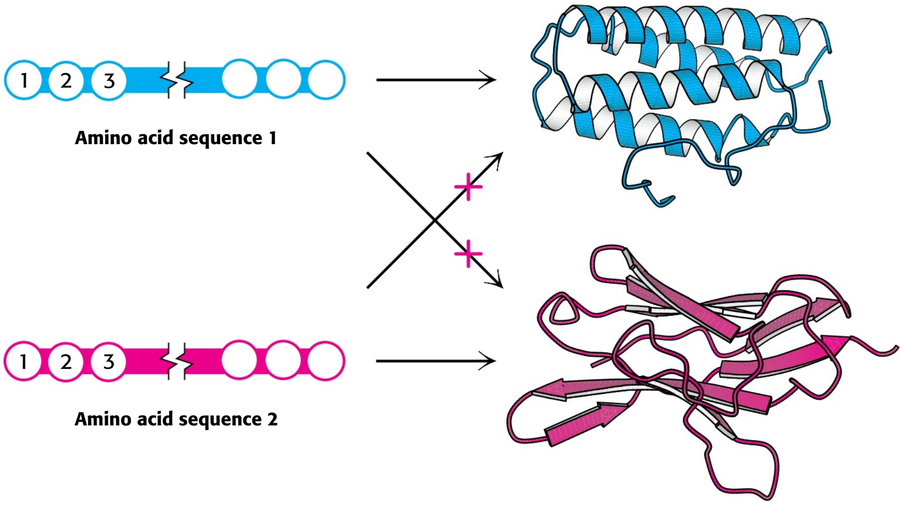


FIGURE 1.23 Protein folding. Proteins are linear polymers of amino acids that fold into elaborate structures. The sequence of amino acids determines the three-dimensional structure. Thus, amino acid sequence 1 gives rise only to a protein with the shape depicted in blue, *not* the shape depicted in red.

The fundamental unit of hereditary information, the *gene,* is becoming increasingly difficult to precisely define as our knowledge of the complexities of genetics and genomics increases. The genes that are simplest to define encode the sequences of proteins. For these protein-encoding genes, a block of DNA bases encodes the amino acid sequence of a specific protein molecule. A set of three bases along the DNA strand, called a *codon,* determines the identity of one amino acid within the protein sequence. The set of rules that links the DNA sequence to the encoded protein sequence is called the *genetic code*. One of the biggest surprises from the sequencing of the human genome is the small number of protein-encoding genes. Before the genome-sequencing project began, the consensus view was that the human genome would include approximately 100,000 protein-encoding genes. The current analysis suggests that the actual number is between 20,000 and 25,000. We shall use an estimate of 21,000 throughout this book. However, additional mechanisms allow many genes to encode more than one protein. For example, the genetic information in some genes is translated in more than one way to produce a set of proteins that differ from one another in parts of their amino acid sequences. In other cases, proteins are modified after they have been synthesized through the addition of accessory chemical groups. Through these indirect mechanisms, much more complexity is encoded in our genomes than would be expected from the number of protein-encoding genes alone.

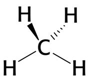
On the basis of current knowledge, the protein-encoding regions account for only about 3% of the human genome. What is the function of the rest of the DNA? Some of it contains information that regulates the expression of specific genes (i.e., the production of specific proteins) in particular cell types and physiological conditions. Essentially every human cell contains the same DNA genome, yet cell types differ considerably in the proteins that they produce. For example, hemoglobin is expressed only in precursors of red blood cells, even though the genes for hemoglobin are present in essentially every cell. Specific sets of genes are expressed in response to hormones, even though these genes are not expressed in the same cell in the absence of the hormones. The control regions that regulate such differences account for only a small amount of the remainder of our genomes. The truth is that we do not yet understand all of the function of much of the remainder of the DNA. Some of it is sometimes referred to as "junk"—stretches of DNA that were inserted at some stage of evolution and have remained. In some cases, this DNA may, in fact, serve important functions. In others, it may serve no function but, because it does not cause significant harm, it has remained.

APPENDIX: Visualizing Molecular Structures I: Small Molecules

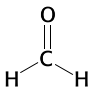
The authors of a biochemistry textbook face the problem of trying to present three-dimensional molecules in the two dimensions available on the printed page. The interplay between the three-dimensional structures of biomolecules and their biological functions will be discussed extensively throughout this book. Toward this end, we will frequently use representations that, although of necessity are rendered in two dimensions, emphasize the three-dimensional structures of molecules.

Stereochemical Renderings

Most of the chemical formulas in this book are drawn to depict the geometric arrangement of atoms, crucial to chemical bonding and reactivity, as accurately as possible. For example, the carbon atom of methane is tetrahedral, with H–C–H angles of 109.5 degrees, whereas the carbon atom in formaldehyde has bond angles of 120 degrees.



Methane

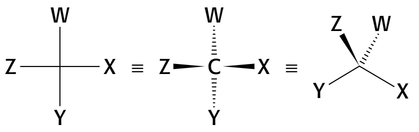


Formaldehyde

To illustrate the correct *stereochemistry* about tetrahedral carbon atoms, wedges will be used to depict the direction of a bond into or out of the plane of the page. A solid wedge with the broad end away from the carbon atom denotes a bond coming toward the viewer out of the plane. A dashed wedge, with its broad end at the carbon atom, represents a bond going away from the viewer behind the plane of the page. The remaining two bonds are depicted as straight lines.

Fischer Projections

Although representative of the actual structure of a compound, stereochemical structures are often difficult to draw quickly. An alternative, less-representative method of depicting structures with tetrahedral carbon centers relies on the use of *Fischer projections*.



Fischer projection

Stereochemical rendering

In a Fischer projection, the bonds to the central carbon are represented by horizontal and vertical lines from the substituent atoms to the carbon atom, which is assumed to be at the center of the cross. By convention, the horizontal bonds are assumed to project out of the page toward the viewer, whereas the vertical bonds are assumed to project behind the page away from the viewer.

Molecular Models for Small Molecules

For depicting the molecular architecture of small molecules in more detail, two types of models will often be used: space filling and ball and stick. These models show structures at the atomic level.

1. *Space-Filling Models*. The space-filling models are the most realistic. The size and position of an atom in a space-filling model are determined by its bonding properties and van der Waals radius, or contact distance. A van der Waals radius describes how closely two atoms can approach each other when they are not linked by a covalent bond. The colors of the model are set by convention.

Carbon, black Hydrogen, white Nitrogen, blue
Oxygen, red Sulfur, yellow Phosphorus, purple

Space-filling models of several simple molecules are shown in Figure 1.24.

2. *Ball-and-Stick Models*. Ball-and-stick models are not as realistic as space-filling models, because the atoms are depicted as spheres of radii smaller than their van der Waals radii. However, the bonding arrangement is easier to see because the bonds are explicitly represented as sticks. In an illustration, the taper of a stick, representing parallax, tells which of a pair of bonded atoms is closer to the reader. A ball-and-stick

model reveals a complex structure more clearly than a space-filling model does. Ball-and-stick models of several simple molecules are shown in Figure 1.24.

Molecular models for depicting large molecules will be discussed in the appendix to Chapter 2.

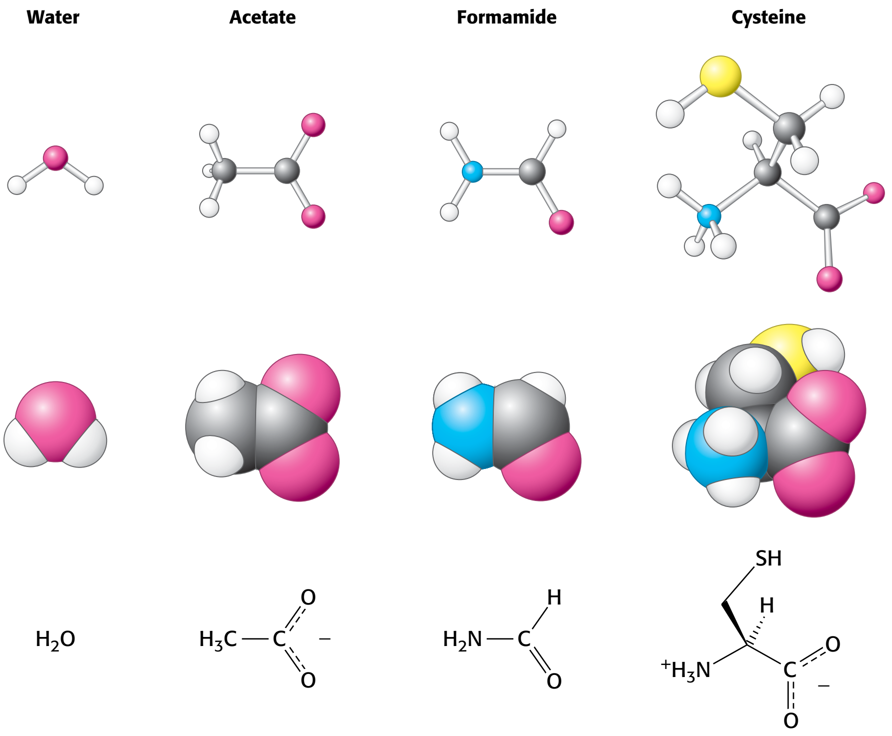


FIGURE 1.24 Molecular representations. Structural formulas (bottom), ball-and-stick models (top), and space-filling representations (middle) of selected molecules are shown. Black = carbon, red = oxygen, white = hydrogen, yellow = sulfur, blue = nitrogen.

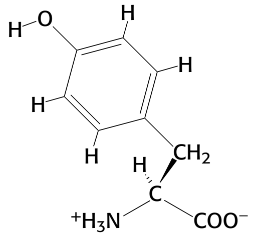
KEY TERMS

biological macromolecule (p. 2)
metabolite (p. 2)
deoxyribonucleic acid (DNA) (p. 2)
protein (p. 2)
Eukarya (p. 3)
Bacteria (p. 3)
Archaea (p. 3)
eukaryote (p. 3)
prokaryote (p. 3)
double helix (p. 5)
covalent bond (p. 5)
resonance structure (p. 7)
ionic interaction (p. 7)
hydrogen bond (p. 8)
van der Waals interaction (p. 8)
hydrophobic effect (p. 9)
hydrophobic interaction (p. 9)
entropy (p. 11)
enthalpy (p. 11)
free energy (Gibbs free energy) (p. 12)
pH (p. 13)
pK_a value (p. 14)
buffer (p. 15)
predisposition (p. 18)
microbiome (p. 19)
amino acid (p. 21)
genetic code (p. 21)

PROBLEMS

1. *Donors and acceptors*. Identify the hydrogen-bond donors and acceptors in each of the four bases on page 4.

2. *Resonance structures*. The structure of an amino acid, tyrosine, is shown here. Draw an alternative resonance structure.



3. *It takes all types*. What types of noncovalent bonds hold together the following solids?

(a) Table salt (NaCl), which contains Na^+ and Cl^- ions.

(b) Graphite (C), which consists of sheets of covalently bonded carbon atoms.

4. *Don't break the law*. Given the following values for the changes in enthalpy (ΔH) and entropy (ΔS), which of the following processes can take place at 298 K without violating the Second Law of Thermodynamics?

(a) $\Delta H = -84\ \text{kJ mol}^{-1}\ (-20\ \text{kcal mol}^{-1})$,
$\Delta S = +125\ \text{J mol}^{-1}\ \text{K}^{-1}\ (+30\ \text{cal mol}^{-1}\ \text{K}^{-1})$

(b) $\Delta H = -84$ kJ mol^{-1} (-20 kcal mol^{-1}),
$\Delta S = -125$ J mol^{-1} K^{-1} (-30 cal mol^{-1} K^{-1})

(c) $\Delta H = +84$ kJ mol^{-1} ($+20$ kcal mol^{-1}),
$\Delta S = +125$ J mol^{-1} K^{-1} ($+30$ cal mol^{-1} K^{-1})

(d) $\Delta H = +84$ kJ mol^{-1} ($+20$ kcal mol^{-1}),
$\Delta S = -125$ J mol^{-1} K^{-1} (-30 cal mol^{-1} K^{-1})

5. *Double-helix-formation entropy*. For double-helix formation, ΔG can be measured to be -54 kJ mol^{-1} (-13 kcal mol^{-1}) at pH 7.0 in 1 M NaCl at 25°C (298 K). The heat released indicates an enthalpy change of -251 kJ mol^{-1} (-60 kcal mol^{-1}). For this process, calculate the entropy change for the system and the entropy change for the surroundings.

6. *Find the pH*. What are the pH values for the following solutions?

(a) 0.1 M HCl

(b) 0.1 M NaOH

(c) 0.05 M HCl

(d) 0.05 M NaOH

7. *A weak acid*. What is the pH of a 0.1 M solution of acetic acid ($pK_a = 4.75$)?

(Hint: Let x be the concentration of H^+ ions released from acetic acid when it dissociates. The solutions to a quadratic equation of the form $ax^2 + bx + c = 0$ are $x = (-b \pm \sqrt{b^2 - 4ac})/2a$.)

8. *Substituent effects*. What is the pH of a 0.1 M solution of chloroacetic acid ($ClCH_2COOH$, $pK_a = 2.86$)?

9. *Water in water*. Given a density of 1 g/ml and a molecular weight of 18 g/mol, calculate the concentration of water in water.

10. *Basic fact*. What is the pH of a 0.1 M solution of ethylamine, given that the pK_a of ethylammonium ion ($CH_3CH_2NH_3^+$) is 10.70?

11. *Comparison*. A solution is prepared by adding 0.01 M acetic acid and 0.01 M ethylamine to water and adjusting the pH to 7.4. What is the ratio of acetate to acetic acid? What is the ratio of ethylamine to ethylammonium ion?

12. *Concentrate*. Acetic acid is added to water until the pH value reaches 4.0. What is the total concentration of the added acetic acid?

13. *Dilution*. 100 mL of a solution of hydrochloric acid with pH 5.0 is diluted to 1 L. What is the pH of the diluted solution?

14. *Buffer dilution*. 100 mL of a 0.1 mM buffer solution made from acetic acid and sodium acetate with pH 5.0 is diluted to 1 L. What is the pH of the diluted solution?

15. *Find the pK_a*. For an acid HA, the concentrations of HA and A^- are 0.075 and 0.025, respectively, at pH 6.0. What is the pK_a value for HA?

16. *pH indicator*. A dye that is an acid and that appears as different colors in its protonated and deprotonated forms can be used as a pH indicator. Suppose that you have a 0.001 M solution of a dye with a pK_a of 7.2. From the color, the concentration of the protonated form is found to be 0.0002 M. Assume that the remainder of the dye is in the deprotonated form. What is the pH of the solution?

17. *What's the ratio?* An acid with a pK_a of 8.0 is present in a solution with a pH of 6.0. What is the ratio of the protonated to the deprotonated form of the acid?

18. *Phosphate buffer*. What is the ratio of the concentrations of $H_2PO_4^-$ and HPO_4^{2-} at (a) pH 7.0; (b) pH 7.5; (c) pH 8.0?

19. *Neutralization of phosphate*. Given that phosphoric acid (H_3PO_4) can give up three protons with different pK_a values, sketch a plot of pH as a function of added drops of sodium hydroxide solution, starting with a solution of phosphoric acid at pH 1.0.

20. *Buffer capacity*. Two solutions of sodium acetate are prepared, one with a concentration of 0.1 M and the other with a concentration of 0.01 M. Calculate the pH values when the following concentrations of HCl have been added to each of these solutions: 0.0025 M, 0.005 M, 0.01 M, and 0.05 M.

21. *Buffer preparation*. You wish to prepare a buffer consisting of acetic acid and sodium acetate with a total acetic acid plus acetate concentration of 250 mM and a pH of 5.0. What concentrations of acetic acid and sodium acetate should you use? Assuming you wish to make 2 liters of this buffer, how many moles of acetic acid and sodium acetate will you need? How many grams of each will you need (molecular weights: acetic acid 60.05 g mol^{-1}, sodium acetate, 82.03 g mol^{-1})?

22. *An alternative approach*. When you go to prepare the buffer described in Problem 21, you discover that your laboratory is out of sodium acetate, but you do have sodium hydroxide. How much (in moles and grams) acetic acid and sodium hydroxide do you need to make the buffer?

23. *Another alternative*. Your friend from another laboratory was out of acetic acid, so tries to prepare the buffer in Problem 21 by dissolving 41.02 g of sodium acetate in water, carefully adding 180.0 ml of 1 M HCl, and adding more water to reach a total volume of 2 liters. What is the total concentration of acetate plus acetic acid in the solution? Will this solution have pH 5.0? Will it be identical with the desired buffer? If not, how will it differ?

24. *Blood substitute*. As noted in this chapter, blood contains a total concentration of phosphate of approximately 1 mM and typically has a pH of 7.4. You wish to make 100 liters of phosphate buffer with a pH of 7.4 from NaH_2PO_4 (molecular weight, 119.98 g mol^{-1}) and Na_2HPO_4 (molecular weight, 141.96 g mol^{-1}). How much of each (in grams) do you need?

25. *A potential problem.* You wish to make a buffer with pH 7.0. You combine 0.060 grams of acetic acid and 14.59 grams of sodium acetate and add water to yield a total volume of 1 liter. What is the pH? Will this be the useful pH 7.0 buffer you seek?

26. *Charge!* Suppose two phosphate groups in DNA (each with a charge of −1) are separated by 12 Å. What is the energy of the ionic interaction between these two phosphates assuming a dielectric constant of 80? Repeat the calculation assuming a dielectric constant of 2.

27. *Vive la différence*. On average, estimate how many base differences there are between two human beings.

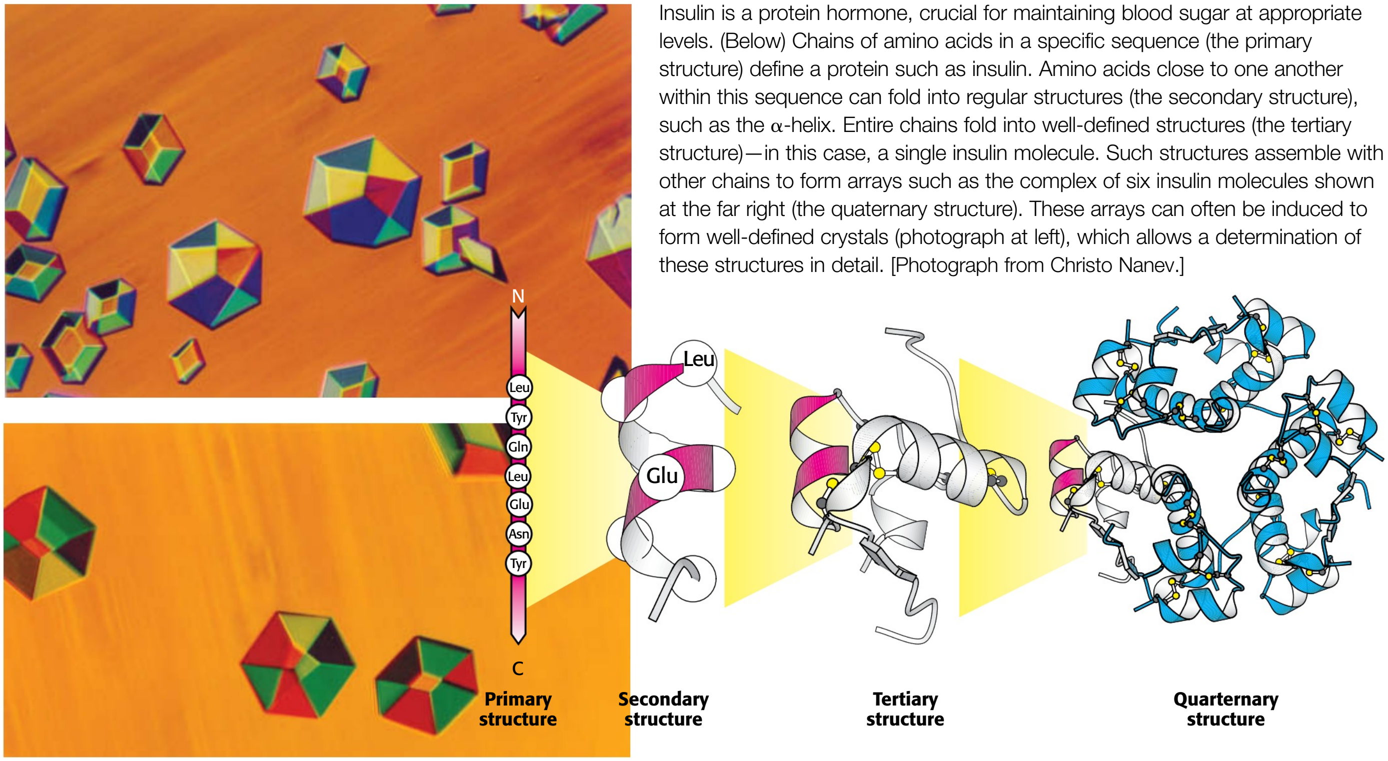
28. *Epigenomics*. The human body contains many distinct cell types yet almost all human cells contain the same genome with 21,000 genes. The distinct cell types are primarily due to differences in gene expression. Assume that one set of 1000 genes is expressed in all cell types and that the remaining 20,000 genes can be divided into sets of 1000 genes that are either all expressed or all not expressed in a given cell type. How many different cell types are possible if each cell type expresses 10 sets of these genes? Note that the number of combinations of n objects into m sets is given by $n!/(m!(n-m)!)$ where $n! = 1*2*\cdots*(n-1)*n$.

29. *Predispositions in populations*. Assume that 10% of the members of a population will get a particular disease over the course of their lifetime. Genomic studies reveal that 5% of the population have sequences in their genomes such that their probability of getting the disease over the course of their lifetimes is 50%. What is the average lifetime risk of this disease for the remaining 95% of the population without these sequences?

CHAPTER

2

Protein Composition and Structure



Insulin is a protein hormone, crucial for maintaining blood sugar at appropriate levels. (Below) Chains of amino acids in a specific sequence (the primary structure) define a protein such as insulin. Amino acids close to one another within this sequence can fold into regular structures (the secondary structure), such as the α-helix. Entire chains fold into well-defined structures (the tertiary structure)—in this case, a single insulin molecule. Such structures assemble with other chains to form arrays such as the complex of six insulin molecules shown at the far right (the quaternary structure). These arrays can often be induced to form well-defined crystals (photograph at left), which allows a determination of these structures in detail. [Photograph from Christo Nanev.]

Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation. Indeed, much of this book will focus on understanding what proteins do and how they perform these functions.

Several key properties enable proteins to participate in a wide range of functions.

1. *Proteins are linear polymers built of monomer units called amino acids,* which are linked end to end. The sequence of linked amino acids is called the primary structure. Remarkably, proteins spontaneously fold up into three-dimensional structures that are determined by the sequence of amino acids in the protein polymer. Three-dimensional structure formed by hydrogen bonds between amino acids near one another is called secondary structure, whereas tertiary structure is formed by long-range interactions between amino acids. Protein function depends directly on this three-dimensional structure (Figure 2.1). Thus, *proteins are the embodiment of the transition from the one-dimensional world of sequences to the three-dimensional world of molecules capable of diverse activities.* Many proteins also display

OUTLINE

2.1 Proteins Are Built from a Repertoire of 20 Amino Acids

2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops

2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores

2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures

2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

quaternary structure, in which the functional protein is composed of several distinct polypeptide chains.

2. *Proteins contain a wide range of functional groups*. These functional groups include alcohols, thiols, thioethers, carboxylic acids, carboxamides, and a variety of basic groups. Most of these groups are chemically reactive. When combined in various sequences, this array of functional groups accounts for the broad spectrum of protein function. For instance, their reactive properties are essential to the function of *enzymes*, the proteins that catalyze specific chemical reactions in biological systems (Chapters 8 through 10).

3. *Proteins can interact with one another and with other biological macromolecules to form complex assemblies*. The proteins within these assemblies can act synergistically to generate capabilities that individual proteins may lack. Examples of these assemblies include macromolecular machines that replicate DNA, transmit signals within cells, and enable muscle cells to contract (Figure 2.2).

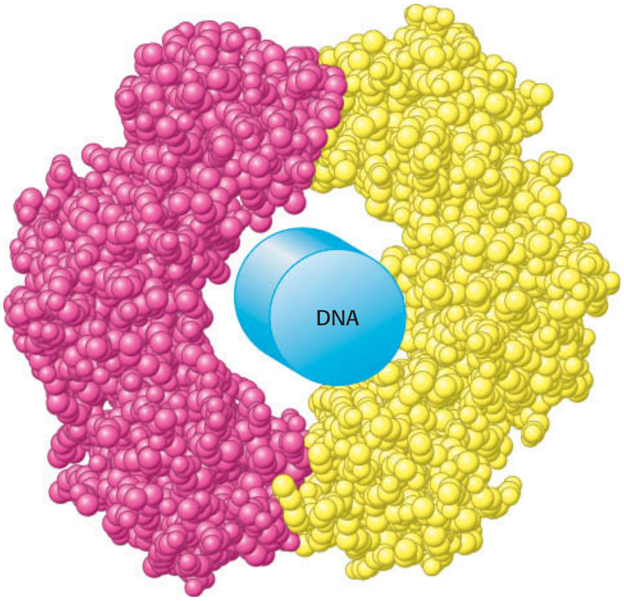


FIGURE 2.1 Structure dictates function. A protein component of the DNA replication machinery surrounds a section of DNA double helix depicted as a cylinder. The protein, which consists of two identical subunits (shown in red and yellow), acts as a clamp that allows large segments of DNA to be copied without the replication machinery dissociating from the DNA. [Drawn from 2POL.pdb.]

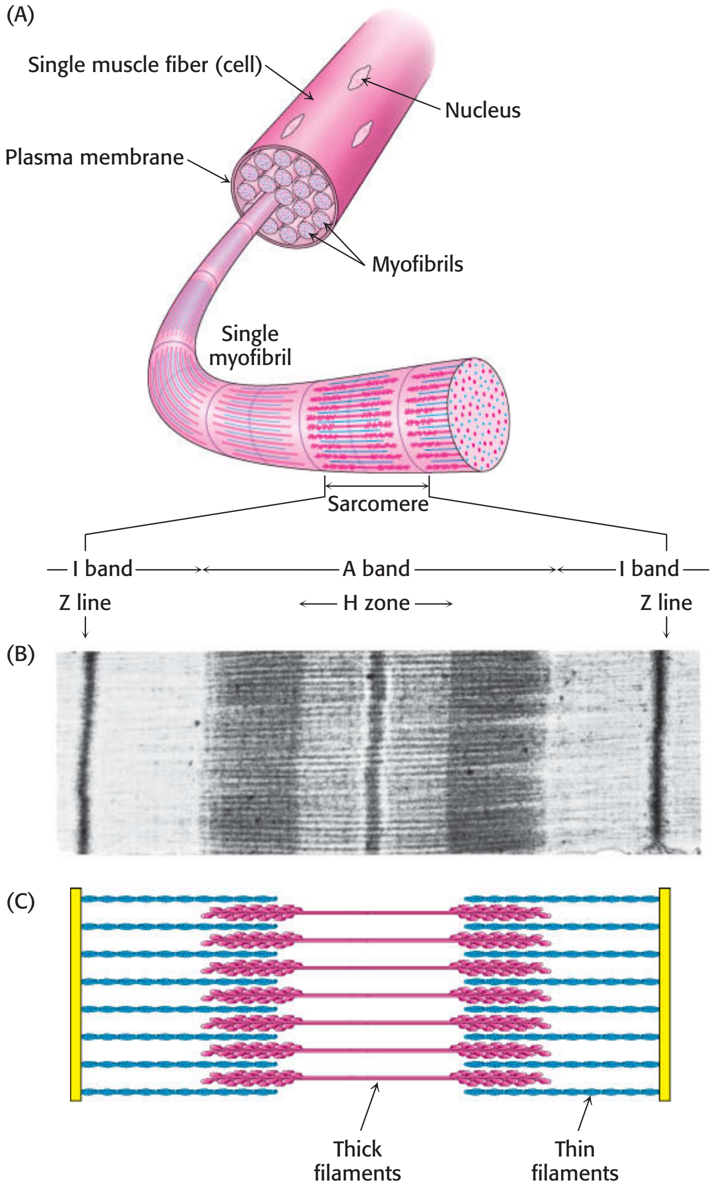


FIGURE 2.2 A complex protein assembly. (A) A single muscle cell contains multiple myofibrils, each of which is comprised of numerous repeats of a complex protein assembly known as the sarcomere. (B) The banding pattern of a sarcomere, evident by electron microscopy, is caused by (C) the interdigitation of filaments made up of many individual proteins. [(B) Courtesy of Dr. Hugh Huxley.]

4. *Some proteins are quite rigid, whereas others display considerable flexibility.* Rigid units can function as structural elements in the cytoskeleton (the internal scaffolding within cells) or in connective tissue. Proteins with some flexibility may act as hinges, springs, or levers. In addition, conformational changes within proteins enable the regulated assembly of larger protein complexes as well as the transmission of information within and between cells (Figure 2.3).

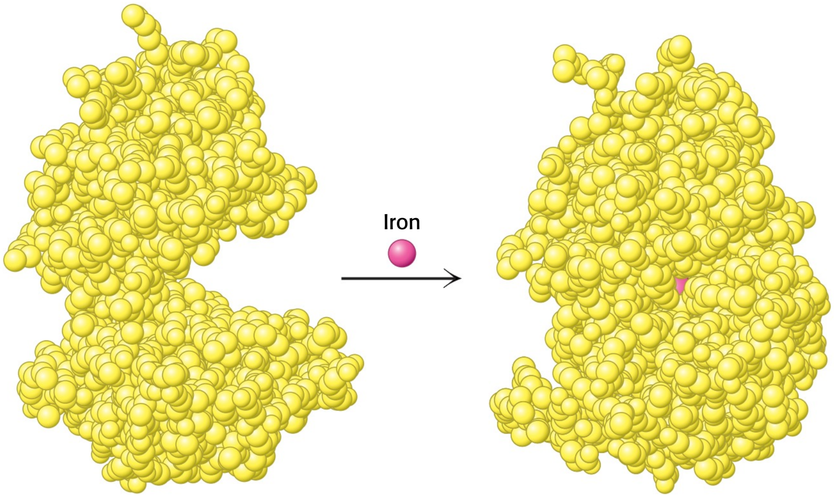


FIGURE 2.3 Flexibility and function. On binding iron, the protein lactoferrin undergoes a substantial change in conformation that allows other molecules to distinguish between the iron-free and the iron-bound forms. [Drawn from 1LFH.pdb and 1LFG.pdb.]

2.1 Proteins Are Built from a Repertoire of 20 Amino Acids

Amino acids are the building blocks of proteins. An *α-amino acid* consists of a central carbon atom, called the *α carbon,* linked to an amino group, a carboxylic acid group, a hydrogen atom, and a distinctive R group. The R group is often referred to as the *side chain*. With four different groups connected to the tetrahedral α-carbon atom, α-amino acids are *chiral:* they may exist in one or the other of two mirror-image forms, called the L isomer and the D isomer (Figure 2.4).

Only L *amino acids are constituents of proteins*. For almost all amino acids, the L isomer has *S* (rather than *R*) absolute configuration (Figure 2.5). What is the basis for the preference for L amino acids? The answer has been lost to evolutionary history. It is possible that the preference for L over D amino acids was a consequence of a chance selection. However, there is evidence that L amino acids are slightly more soluble than a racemic mixture of D and L amino acids, which tend to form crystals. This small solubility difference could have been amplified over time so that the L isomer became dominant in solution.

Amino acids in solution at neutral pH exist predominantly as *dipolar ions* (also called *zwitterions*). In the dipolar form, the amino group is protonated

Notation for distinguishing stereoisomers

The four different substituents of an asymmetric carbon atom are assigned a priority according to atomic number. The lowest-priority substituent, often hydrogen, is pointed away from the viewer. The configuration about the carbon atom is called *S* (from the Latin *sinister*, "left") if the progression from the highest to the lowest priority is counterclockwise. The configuration is called *R* (from the Latin *rectus*, "right") if the progression is clockwise.

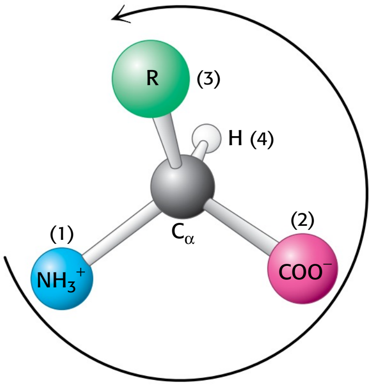


FIGURE 2.5 Only L amino acids are found in proteins. Almost all L amino acids have an *S* absolute configuration. The counterclockwise direction of the arrow from highest- to lowest-priority substituents indicates that the chiral center is of the *S* configuration.

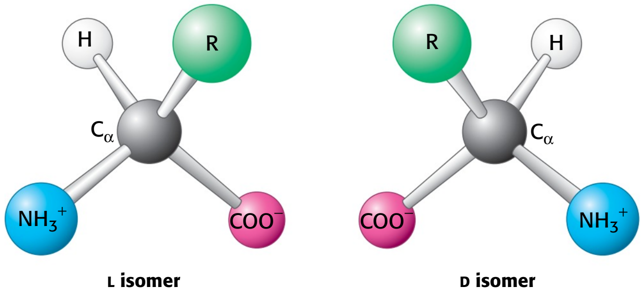


FIGURE 2.4 The L and D isomers of amino acids. The letter R refers to the side chain. The L and D isomers are mirror images of each other.

(—NH_3^+) and the carboxyl group is deprotonated (—COO^-). The ionization state of an amino acid varies with pH (Figure 2.6). In acid solution (e.g., pH 1), the amino group is protonated (—NH_3^+) and the carboxyl group is not dissociated (—COOH). As the pH is raised, the carboxylic acid is the first group to give up a proton, inasmuch as its pK_a is near 2. The dipolar form persists until the pH approaches 9, when the protonated amino group loses a proton.

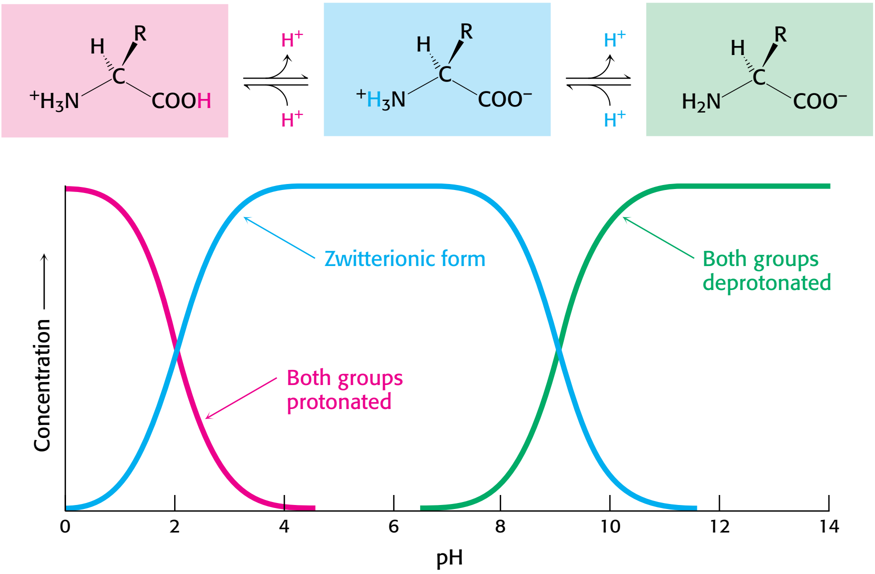


FIGURE 2.6 Ionization state as a function of pH. The ionization state of amino acids is altered by a change in pH. The zwitterionic form predominates near physiological pH.

Twenty kinds of side chains varying in *size, shape, charge, hydrogen-bonding capacity, hydrophobic character,* and *chemical reactivity* are commonly found in proteins. Indeed, all proteins in all species—bacterial, archaeal, and eukaryotic—are constructed from the same set of 20 amino acids with only a few exceptions. This fundamental alphabet for the construction of proteins is several billion years old. The remarkable range of functions mediated by proteins results from the diversity and versatility of these 20 building blocks. Understanding how this alphabet is used to create the intricate three-dimensional structures that enable proteins to carry out so many biological processes is an exciting area of biochemistry and one that we will return to in Section 2.6.

Although there are many ways to classify amino acids, we will sort these molecules into four groups, on the basis of the general chemical characteristics of their R groups:

1. Hydrophobic amino acids with nonpolar R groups
2. Polar amino acids with neutral R groups but the charge is not evenly distributed
3. Positively charged amino acids with R groups that have a positive charge at physiological pH
4. Negatively charged amino acids with R groups that have a negative charge at physiological pH

Hydrophobic amino acids. The simplest amino acid is *glycine,* which has a single hydrogen atom as its side chain. With two hydrogen atoms bonded to the α-carbon atom, glycine is unique in being *achiral. Alanine,* the next simplest amino acid, has a methyl group (—CH_3) as its side chain (Figure 2.7).

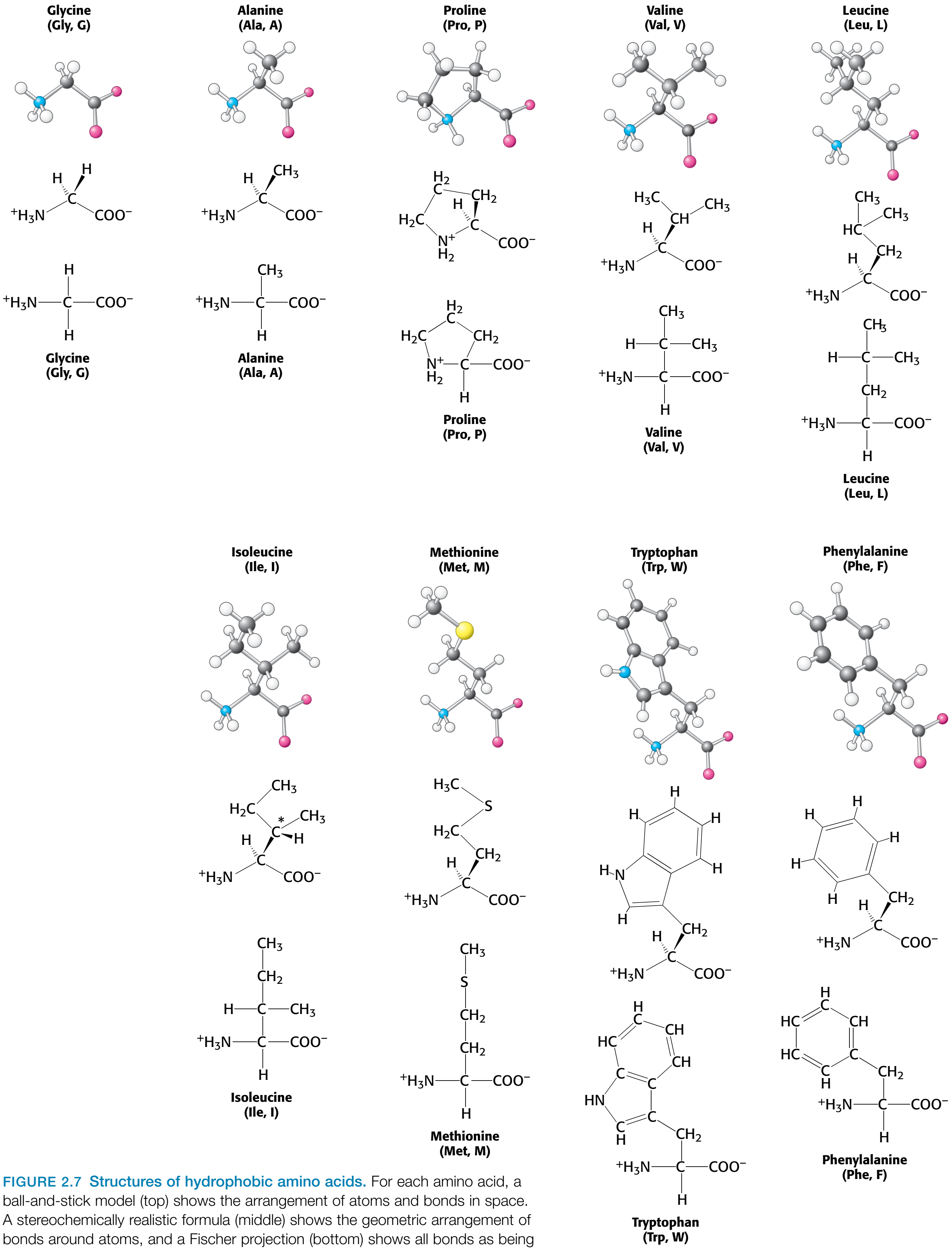

Glycine
(Gly, G)
Alanine
(Ala, A)
Proline
(Pro, P)
Valine
(Val, V)
Leucine
(Leu, L)
Isoleucine
(Ile, I)
Methionine
(Met, M)
Tryptophan
(Trp, W)
Phenylalanine
(Phe, F)

FIGURE 2.7 Structures of hydrophobic amino acids. For each amino acid, a ball-and-stick model (top) shows the arrangement of atoms and bonds in space. A stereochemically realistic formula (middle) shows the geometric arrangement of bonds around atoms, and a Fischer projection (bottom) shows all bonds as being perpendicular for a simplified representation (see the Appendix to Chapter 1). The additional chiral center in isoleucine is indicated by an asterisk.

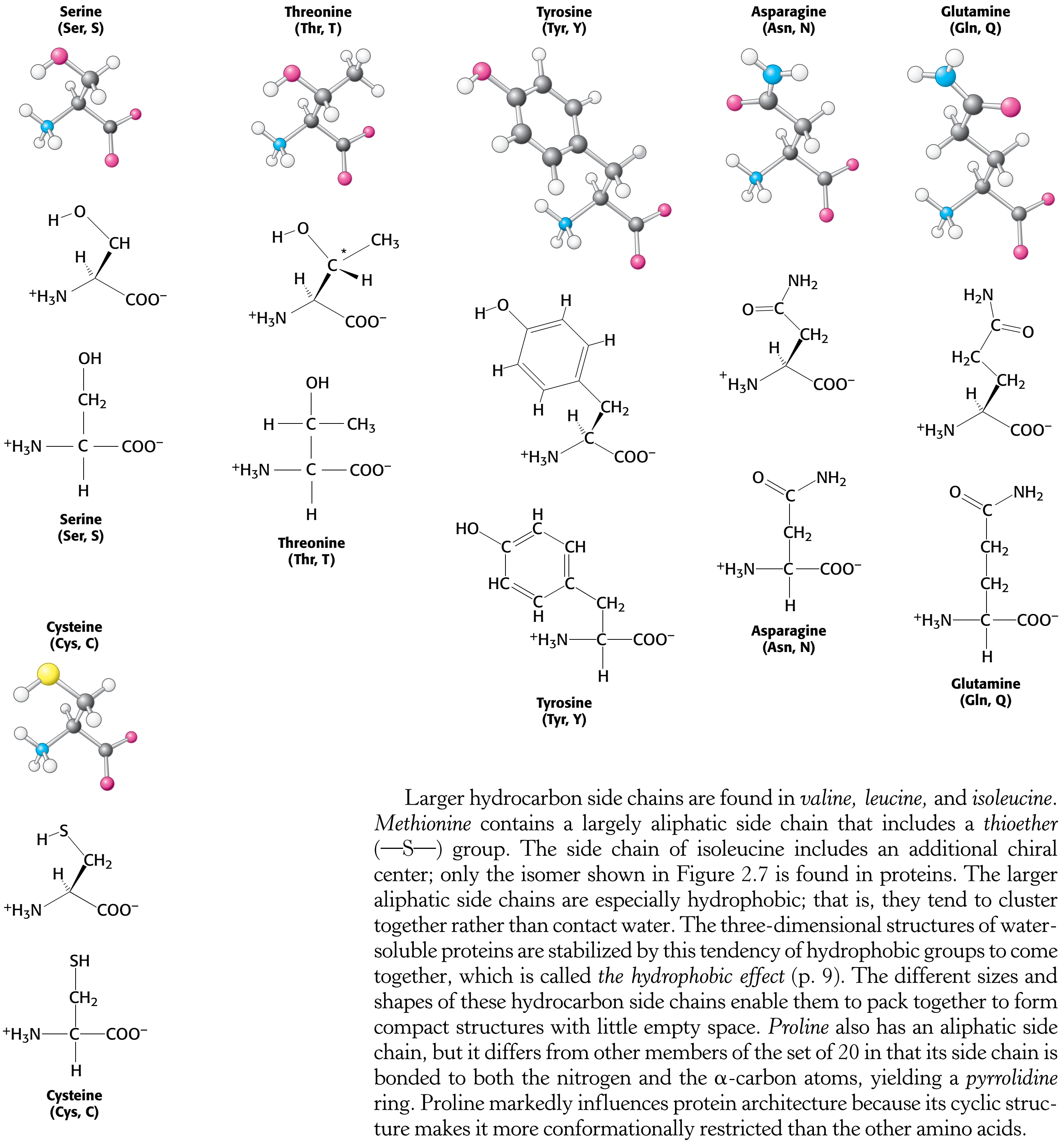
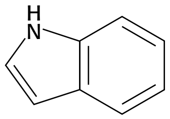


FIGURE 2.8 **Structures of the polar amino acids.** The additional chiral center in threonine is indicated by an asterisk.

Larger hydrocarbon side chains are found in *valine, leucine,* and *isoleucine*. *Methionine* contains a largely aliphatic side chain that includes a *thioether* (—S—) group. The side chain of isoleucine includes an additional chiral center; only the isomer shown in Figure 2.7 is found in proteins. The larger aliphatic side chains are especially hydrophobic; that is, they tend to cluster together rather than contact water. The three-dimensional structures of water-soluble proteins are stabilized by this tendency of hydrophobic groups to come together, which is called *the hydrophobic effect* (p. 9). The different sizes and shapes of these hydrocarbon side chains enable them to pack together to form compact structures with little empty space. *Proline* also has an aliphatic side chain, but it differs from other members of the set of 20 in that its side chain is bonded to both the nitrogen and the α-carbon atoms, yielding a *pyrrolidine* ring. Proline markedly influences protein architecture because its cyclic structure makes it more conformationally restricted than the other amino acids.

Two amino acids with relatively simple *aromatic side chains* are part of the fundamental repertoire. *Phenylalanine,* as its name indicates, contains a phenyl ring attached in place of one of the hydrogen atoms of alanine. *Tryptophan* has an *indole* group joined to a methylene (—CH_2—) group; the indole group comprises two fused rings containing an NH group. Phenylalanine is purely hydrophobic, whereas tryptophan is less so because of its NH group.



Indole

Polar amino acids. Six amino acids are polar but uncharged. Three amino acids, *serine, threonine,* and *tyrosine,* contain *hydroxyl groups* (—OH) attached to a hydrophobic side chain (Figure 2.8). Serine can be thought of as a version of alanine with a hydroxyl group attached, threonine resembles valine with a hydroxyl group in place of one of valine's methyl groups, and tyrosine is a version of phenylalanine with the hydroxyl group replacing a hydrogen atom on the aromatic ring. The hydroxyl group makes these amino acids much more *hydrophilic* (water loving) and *reactive* than their hydrophobic analogs. Threonine, like isoleucine, contains an additional asymmetric center; again, only one isomer is present in proteins.

In addition, the set includes *asparagine* and *glutamine*, two amino acids that contain a terminal *carboxamide*. The side chain of glutamine is one methylene group longer than that of asparagine.

Cysteine is structurally similar to serine but contains a *sulfhydryl*, or *thiol* (—SH), group in place of the hydroxyl (—OH) group. The sulfhydryl group is much more reactive. Pairs of sulfhydryl groups may come together to form disulfide bonds, which are particularly important in stabilizing some proteins, as will be discussed shortly.

Positively charged amino acids. We turn now to amino acids with complete positive charges that render them highly hydrophilic. *Lysine* and *arginine* have long side chains that terminate with groups that are *positively charged* at neutral pH. Lysine is capped by a primary amino group and arginine by a guanidinium group. *Histidine* contains an imidazole group, an aromatic ring that also can be positively charged (Figure 2.9).

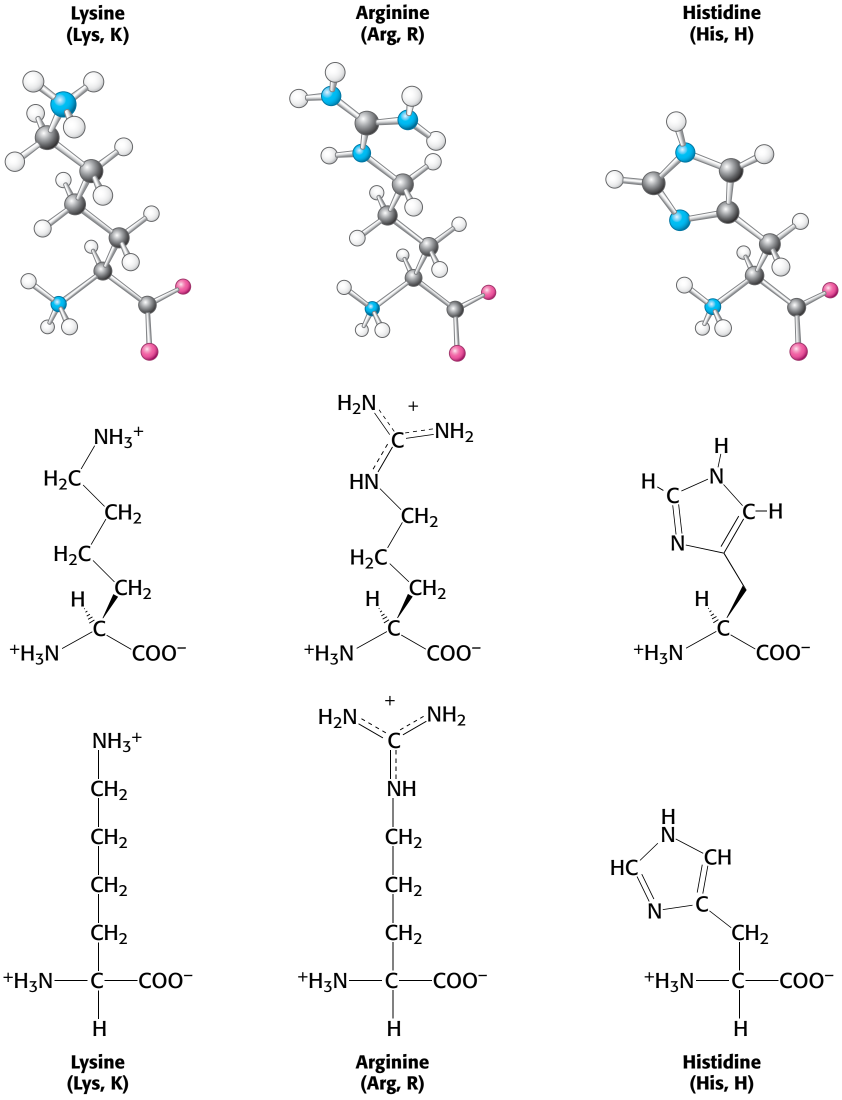
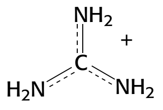
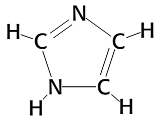


FIGURE 2.9 Positively charged amino acids lysine, arginine, and histidine.

With a pK_a value near 6, the imidazole group can be uncharged or positively charged near neutral pH, depending on its local environment (Figure 2.10). Histidine is often found in the active sites of enzymes, where the imidazole ring can bind and release protons in the course of enzymatic reactions.



Guanidinium



Imidazole

Negatively charged amino acids. This set of amino acids contains two with *acidic side chains: aspartic acid* and *glutamic acid* (Figure 2.11). These amino acids are charged derivatives of asparagine and glutamine (Figure 2.8), with a carboxylic acid in place of a carboxamide. Aspartic acid and glutamic acid are often called *aspartate* and *glutamate* to emphasize that, at physiological pH, their side chains usually lack a proton that is present in the acid form and hence are negatively charged. Nonetheless, these side chains can accept protons in some proteins, often with functionally important consequences.

Seven of the 20 amino acids have readily ionizable side chains. These 7 amino acids are able to donate or accept protons to facilitate reactions as well as to form ionic bonds. Table 2.1 gives equilibria and typical pK_a values for ionization of the side chains of tyrosine, cysteine, arginine, lysine,

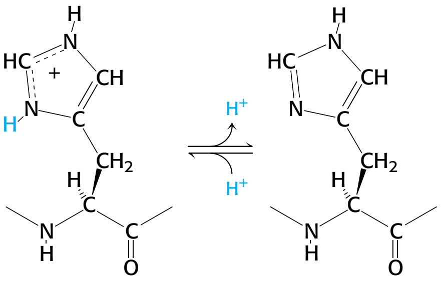


FIGURE 2.10 Histidine ionization. Histidine can bind or release protons near physiological pH.

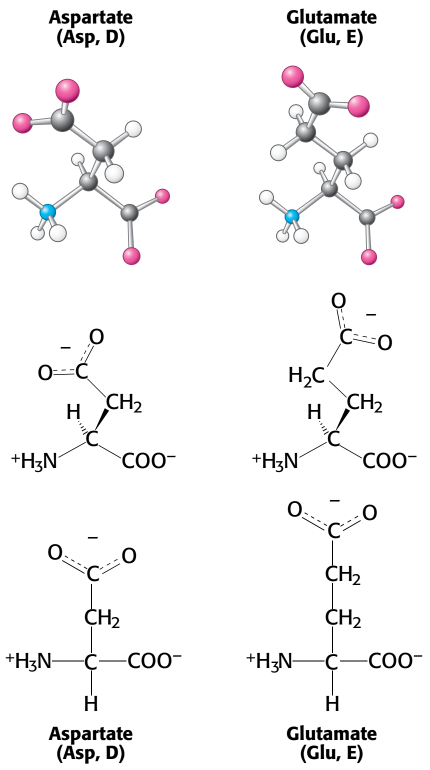
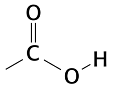
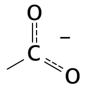
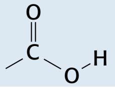
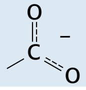
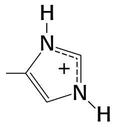
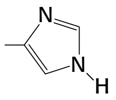
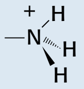
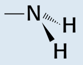
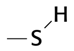
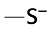
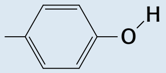
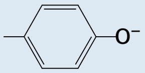
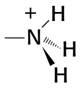
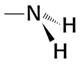
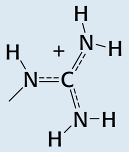
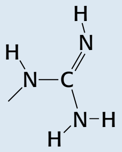


FIGURE 2.11 Negatively charged amino acids.

TABLE 2.1 Typical pK_a values of ionizable groups in proteins

Group	Acid	⇌	Base	Typical pK_a*
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0
Cysteine		⇌		8.3
Tyrosine		⇌		10.9
Lysine		⇌		10.8
Arginine		⇌		12.5

*pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group

histidine, and aspartic and glutamic acids in proteins. Two other groups in proteins—the terminal α-amino group and the terminal α-carboxyl group—can be ionized, and typical pK_a values for these groups also are included in Table 2.1.

Amino acids are often designated by either a three-letter abbreviation or a one-letter symbol (Table 2.2). The abbreviations for amino acids are the first three letters of their names, except for asparagine (Asn), glutamine (Gln), isoleucine (Ile), and tryptophan (Trp). The symbols for many amino acids are the first letters of their names (e.g., G for glycine and L for leucine);

TABLE 2.2 Abbreviations for amino acids

Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glutamic acid	Glu	E	Tyrosine	Tyr	Y
Glycine	Gly	G	Valine	Val	V
Histidine	His	H			
Isoleucine	Ile	I	Asparagine or aspartic acid	Asx	B
Leucine	Leu	L			
Lysine	Lys	K	Glutamine or glutamic acid	Glx	Z

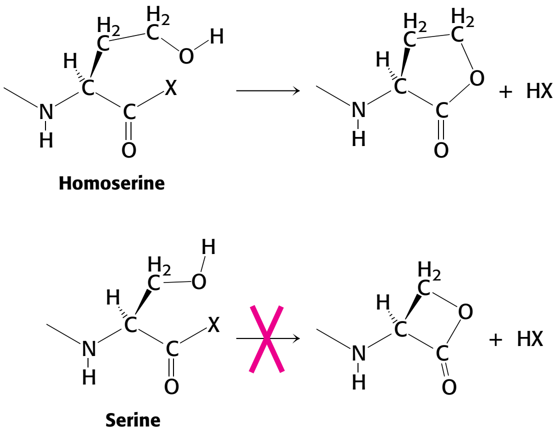


FIGURE 2.12 Undesirable reactivity in amino acids. Some amino acids are unsuitable for proteins because of undesirable cyclization. Homoserine can cyclize to form a stable, five-membered ring, potentially resulting in peptide-bond cleavage. The cyclization of serine would form a strained, four-membered ring and is thus disfavored. X can be an amino group from a neighboring amino acid or another potential leaving group.

the other symbols have been agreed on by convention. These abbreviations and symbols are an integral part of the vocabulary of biochemists.

How did this particular set of amino acids become the building blocks of proteins? First, as a set, they are diverse: their structural and chemical properties span a wide range, endowing proteins with the versatility to assume many functional roles. Second, many of these amino acids were probably available from prebiotic reactions; that is, from reactions that took place before the origin of life. Finally, other possible amino acids may have simply been too reactive. For example, amino acids such as homoserine and homocysteine tend to form five-membered cyclic forms that limit their use in proteins; the alternative amino acids that are found in proteins—serine and cysteine—do not readily cyclize, because the rings in their cyclic forms are too small (Figure 2.12).

2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

Proteins are *linear polymers* formed by linking the α-carboxyl group of one amino acid to the α-amino group of another amino acid. This type of linkage is called a *peptide bond* or an *amide bond*. The formation of a dipeptide from two amino acids is accompanied by the loss of a water molecule (Figure 2.13). The equilibrium of this reaction lies on the side of hydrolysis rather than synthesis under most conditions. Hence, the biosynthesis of peptide bonds requires an input of free energy. Nonetheless, peptide bonds are quite *stable kinetically* because the rate of hydrolysis is extremely slow; the lifetime of a peptide bond in aqueous solution in the absence of a catalyst approaches 1000 years.

A series of amino acids joined by peptide bonds form a *polypeptide chain*, and each amino acid unit in a polypeptide is called a *residue*. *A polypeptide chain has directionality because its ends are different*: an α-amino group

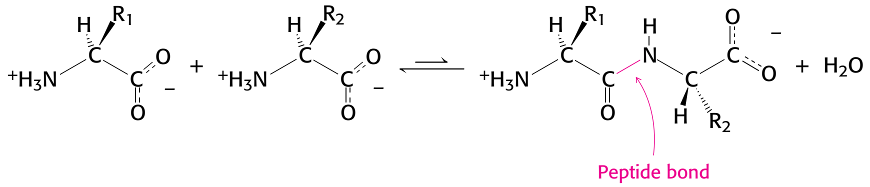


FIGURE 2.13 Peptide-bond formation. The linking of two amino acids is accompanied by the loss of a molecule of water.

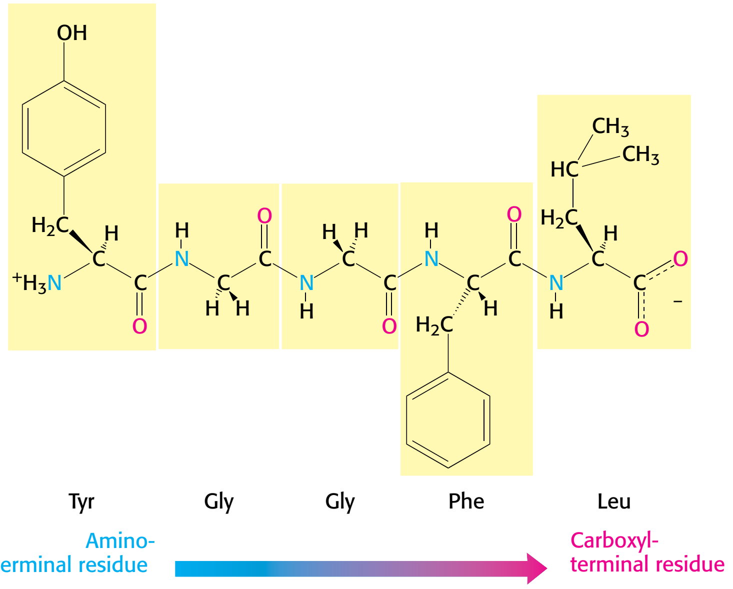


FIGURE 2.14 Amino acid sequences have direction. This illustration of the pentapeptide Tyr-Gly-Gly-Phe-Leu (YGGFL) shows the sequence from the amino terminus to the carboxyl terminus. This pentapeptide, Leu-enkephalin, is an opioid peptide that modulates the perception of pain. The reverse pentapeptide, Leu-Phe-Gly-Gly-Tyr (LFGGY), is a different molecule and has no such effects.

is present at one end and an α-carboxyl group at the other. *The amino end is taken to be the beginning of a polypeptide chain;* by convention, the sequence of amino acids in a polypeptide chain is written starting with the amino-terminal residue. Thus, in the polypeptide Tyr-Gly-Gly-Phe-Leu (YGGFL), tyrosine is the amino-terminal (N-terminal) residue and leucine is the carboxyl-terminal (C-terminal) residue (Figure 2.14). Leu-Phe-Gly-Gly-Tyr (LFGGY) is a different polypeptide, with different chemical properties.

A polypeptide chain consists of a regularly repeating part, called the *main chain* or *backbone,* and a variable part, comprising the distinctive *side chains* (Figure 2.15). The polypeptide backbone is rich in hydrogen-bonding potential. Each residue contains a carbonyl group ($C{=}O$), which is a good hydrogen-bond acceptor, and, with the exception of proline, an NH group, which is a good hydrogen-bond donor. These groups interact with each other and with functional groups from side chains to stabilize particular structures, as will be discussed in Section 2.3.

Dalton
A unit of mass very nearly equal to that of a hydrogen atom. Named after John Dalton (1766–1844), who developed the atomic theory of matter.

Kilodalton (kDa)
A unit of mass equal to 1000 daltons

Most natural polypeptide chains contain between 50 and 2000 amino acid residues and are commonly referred to as *proteins*. The largest single polypeptide known is the muscle protein *titin,* which consists of more than 27,000 amino acids. Polypeptide chains made of small numbers of amino acids are called *oligopeptides* or simply *peptides*. The mean molecular weight of an amino acid residue is about 110 g mol^{-1}, and so the molecular weights of most proteins are between 5500 and 220,000 g mol^{-1}. We can also refer to the mass of a protein, which is expressed in units of daltons; one *dalton* is equal to one atomic mass unit. A protein with a molecular weight of 50,000 g mol^{-1} has a mass of 50,000 daltons, or 50 kDa (kilodaltons).

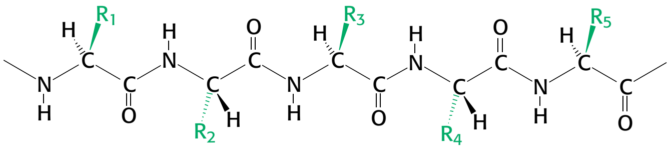


FIGURE 2.15 Components of a polypeptide chain. A polypeptide chain consists of a constant backbone (shown in black) and variable side chains (shown in green).

In some proteins, the linear polypeptide chain is cross-linked. The most common cross-links are *disulfide bonds,* formed by the oxidation of a pair of cysteine residues (Figure 2.16). The resulting unit of two linked cysteines is called *cystine*. Extracellular proteins often have several disulfide bonds, whereas intracellular proteins usually lack them. Rarely, nondisulfide cross-links derived from other side chains are present in proteins. For example, collagen fibers in connective tissue are strengthened in this way, as are fibrin blood clots (Section 10.4).

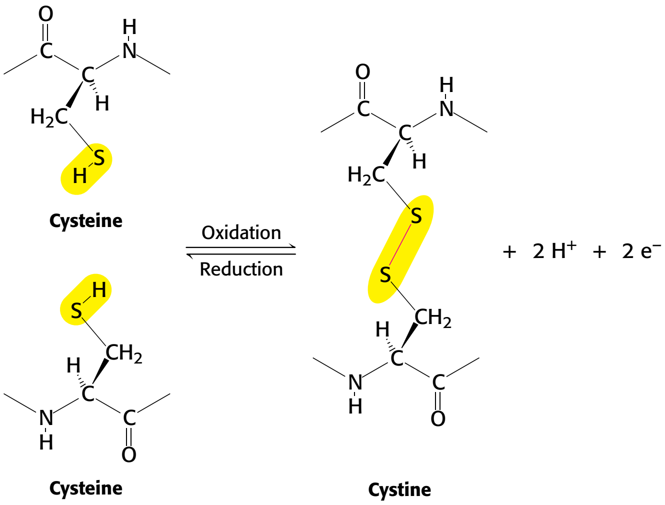


FIGURE 2.16 Cross-links. The formation of a disulfide bond from two cysteine residues is an oxidation reaction.

Proteins have unique amino acid sequences specified by genes

In 1953, Frederick Sanger determined the amino acid sequence of insulin, a protein hormone (Figure 2.17). *This work is a landmark in biochemistry because it showed for the first time that a protein has a precisely defined amino acid sequence* consisting only of L amino acids linked by peptide bonds. This accomplishment stimulated other scientists to carry out sequence studies of a wide variety of proteins. Currently, the complete amino acid sequences of more than 2,000,000 proteins are known. *The striking fact is that each protein has a unique, precisely defined amino acid sequence.* The amino acid sequence of a protein is referred to as its *primary structure.*

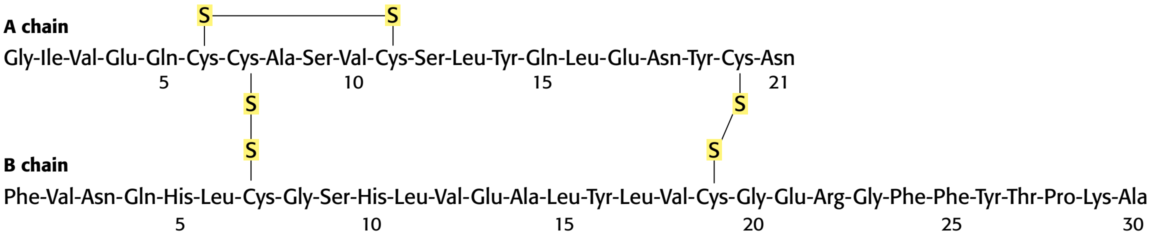

A chain
Gly-Ile-Val-Glu-Gln-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn
5 10 15 21
S S S S S S
B chain
Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala
5 10 15 20 25 30

FIGURE 2.17 Amino acid sequence of bovine insulin.

A series of incisive studies in the late 1950s and early 1960s revealed that the amino acid sequences of proteins are determined by the nucleotide sequences of genes. The sequence of nucleotides in DNA specifies a complementary sequence of nucleotides in RNA, which in turn specifies the amino acid sequence of a protein. In particular, each of the 20 amino acids of the repertoire is encoded by one or more specific sequences of three nucleotides (Section 4.6).

Knowing amino acid sequences is important for several reasons. First, knowledge of the sequence of a protein is usually essential to elucidating its function (e.g., the catalytic mechanism of an enzyme). In fact, proteins with novel properties can be generated by varying the sequence of known proteins. Second, amino acid sequences determine the three-dimensional structures of proteins. The amino acid sequence is the link between the genetic message in DNA and the three-dimensional structure that performs a protein's biological function. Analyses of relations between amino acid sequences and three-dimensional structures of proteins are uncovering the rules that govern the folding of polypeptide chains. Third, alterations in amino acid sequence can lead to abnormal protein function and disease. Severe and sometimes fatal diseases, such as sickle-cell anemia (Chapter 7) and cystic fibrosis, can result from a change in a single amino acid within a protein. Fourth, the sequence of a protein reveals much about its evolutionary history (Chapter 6). Proteins resemble one another in amino acid sequence only if they have a common ancestor. Consequently, molecular events in evolution can be traced from amino acid sequences; molecular paleontology is a flourishing area of research.

Polypeptide chains are flexible yet conformationally restricted

Examination of the geometry of the protein backbone reveals several important features. First, *the peptide bond is essentially planar* (Figure 2.18). Thus, for a pair of amino acids linked by a peptide bond, six atoms lie in the same plane: the α-carbon atom and CO group of the first amino acid and the NH group and α-carbon atom of the second amino acid. The nature of the chemical bonding within a peptide accounts for the bond's planarity. The bond resonates between a single bond and a double bond. Because of this *partial double-bond character,* rotation about this bond is prevented and thus the conformation of the peptide backbone is constrained.

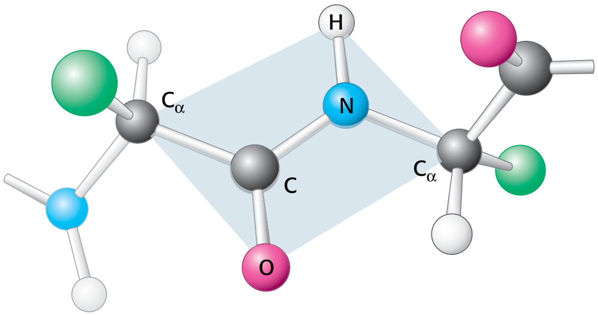
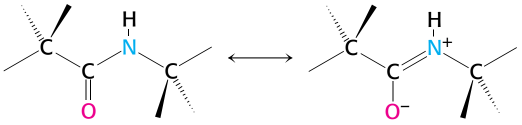


FIGURE 2.18 Peptide bonds are planar. In a pair of linked amino acids, six atoms (C_α, C, O, N, H, and C_α) lie in a plane. Side chains are shown as green balls.



Peptide-bond resonance structures

The partial double-bond character is also expressed in the length of the bond between the CO and the NH groups. As shown in Figure 2.19, the C—N distance in a peptide bond is typically 1.32 Å, which is between the values expected for a C—N single bond (1.49 Å) and a C═N double bond (1.27 Å). Finally, the peptide bond is uncharged, allowing polymers of amino acids linked by peptide bonds to form tightly packed globular structures.

Two configurations are possible for a planar peptide bond. In the *trans* configuration, the two α-carbon atoms are on opposite sides of the peptide bond. In the *cis* configuration, these groups are on the same side of the peptide bond. *Almost all peptide bonds in proteins are trans*. This preference for trans over cis can be explained by the fact that steric clashes between groups attached to the α-carbon atoms hinder the formation of the cis configuration but do not arise in the trans configuration (Figure 2.20). By far the most common cis peptide bonds are X—Pro linkages. Such bonds show less preference for the trans configuration because the nitrogen of proline is bonded to two tetrahedral carbon atoms, limiting the steric differences between the trans and cis forms (Figure 2.21).

In contrast with the peptide bond, the bonds between the amino group and the α-carbon atom and between the α-carbon atom and the carbonyl group are pure single bonds. The two adjacent rigid peptide units can rotate about these bonds, taking on various orientations. *This freedom of rotation about two bonds of each amino acid allows proteins to fold in many different ways*. The rotations about these bonds can be specified by

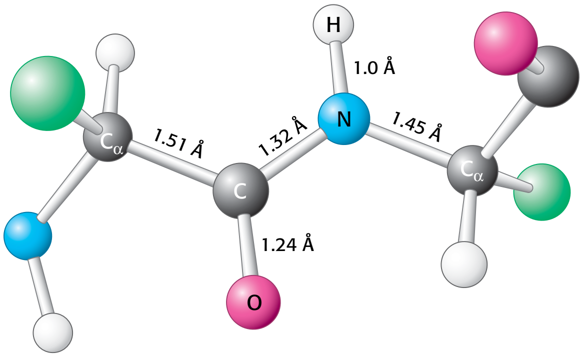


FIGURE 2.19 Typical bond lengths within a peptide unit. The peptide unit is shown in the trans configuration.

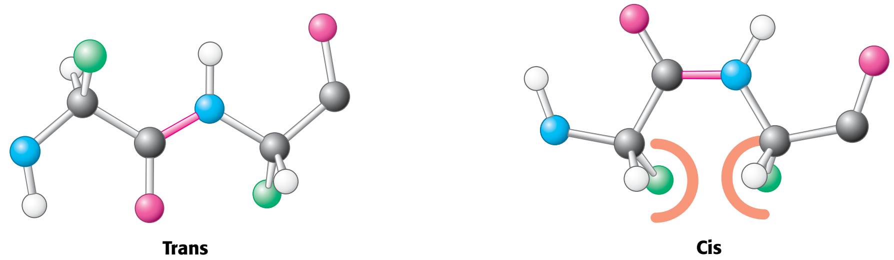


FIGURE 2.20 Trans and cis peptide bonds. The trans form is strongly favored because of steric clashes, indicated by the orange semicircles, that arise in the cis form.

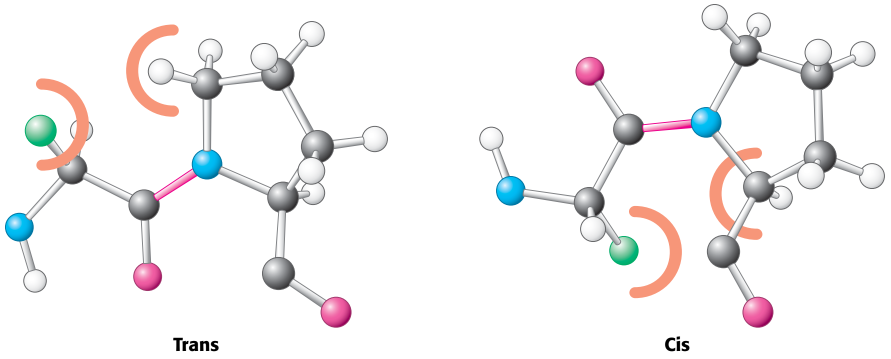


FIGURE 2.21 Trans and cis X–Pro bonds. The energies of these forms are similar to one another because steric clashes, indicated by the orange semicircles, arise in both forms.

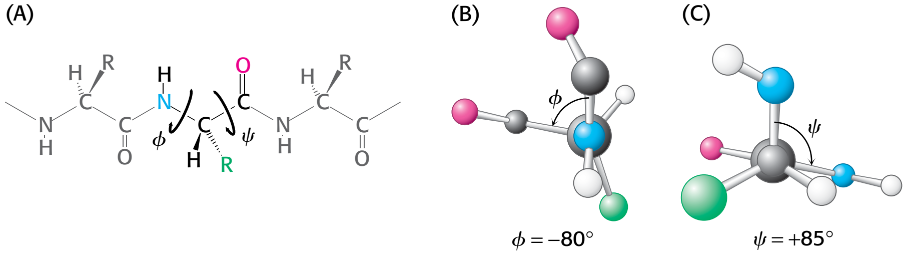


FIGURE 2.22 Rotation about bonds in a polypeptide. The structure of each amino acid in a polypeptide can be adjusted by rotation about two single bonds. (A) Phi (ϕ) is the angle of rotation about the bond between the nitrogen and the α-carbon atoms, whereas psi (ψ) is the angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms. (B) A view down the bond between the nitrogen and the α-carbon atoms, showing how ϕ is measured. (C) A view down the bond between the α-carbon and the carbonyl carbon atoms, showing how ψ is measured.

torsion angles (Figure 2.22). The angle of rotation about the bond between the nitrogen and the α-carbon atoms is called phi (ϕ). The angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms is called psi (ψ). A clockwise rotation about either bond as viewed from the nitrogen atom toward the α-carbon atom or from the α-carbon atom toward the carbonyl group corresponds to a positive value. The ϕ and ψ angles determine the path of the polypeptide chain.

Are all combinations of ϕ and ψ possible? Gopalasamudram Ramachandran recognized that many combinations are forbidden because of steric collisions between atoms. The allowed values can be visualized on a two-dimensional plot called a *Ramachandran plot* (Figure 2.23). Three-quarters of the possible (ϕ, ψ) combinations are excluded simply by local steric clashes. *Steric exclusion, the fact that two atoms cannot be in the same place at the same time, can be a powerful organizing principle.*

Torsion angle
A measure of the rotation about a bond, usually taken to lie between −180 and +180 degrees. Torsion angles are sometimes called dihedral angles.

The ability of biological polymers such as proteins to fold into well-defined structures is remarkable thermodynamically. An unfolded polymer exists as a random coil: each copy of an unfolded polymer will have a different conformation, yielding a mixture of many possible conformations. The favorable entropy associated with a mixture of many conformations opposes folding and must be overcome by interactions favoring the folded form. Thus, highly flexible polymers with a large number of possible conformations do not fold into unique structures. *The rigidity of the peptide unit and the restricted set of allowed ϕ and ψ angles limits the number of structures accessible to the unfolded form sufficiently to allow protein folding to take place.*

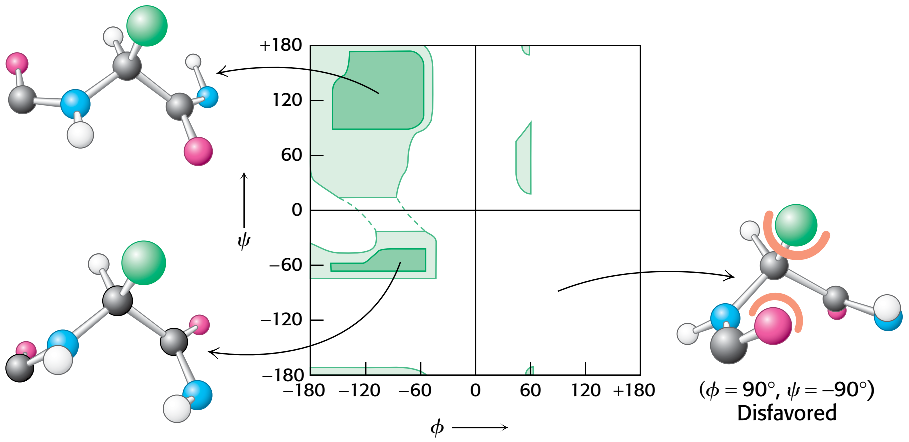


FIGURE 2.23 A Ramachandran plot showing the values of ϕ and ψ. Not all ϕ and ψ values are possible without collisions between atoms. The most favorable regions are shown in dark green; borderline regions are shown in light green. The structure on the right is disfavored because of steric clashes.

2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops

Can a polypeptide chain fold into a regularly repeating structure? In 1951, Linus Pauling and Robert Corey proposed two periodic structures called the *α helix* (alpha helix) and the *β pleated sheet* (beta pleated sheet). Subsequently, other structures such as the *β turn* and *omega* (Ω) *loop* were identified. Although not periodic, these common turn or loop structures are well defined and contribute with α helices and β sheets to form the final protein structure. Alpha helices, β strands, and turns are formed by a regular pattern of hydrogen bonds between the peptide N—H and C═O groups of amino acids that are *near one another in the linear sequence*. Such folded segments are called *secondary structure*.

The alpha helix is a coiled structure stabilized by intrachain hydrogen bonds

In evaluating potential structures, Pauling and Corey considered which conformations of peptides were sterically allowed and which most fully exploited the hydrogen-bonding capacity of the backbone NH and CO groups. The first of their proposed structures, the *α helix*, is a rodlike structure (Figure 2.24). A tightly coiled backbone forms the inner part of the rod and the side chains extend outward in a helical array. The α helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. In particular, the CO group of each amino acid forms a hydrogen bond with the NH group of the amino acid that is situated four residues ahead in the sequence (Figure 2.25). Thus, except for amino acids near the ends of an α helix, *all the main-chain CO and NH groups are hydrogen bonded.* Each residue is related to the next one by a rise, also called *translation,* of 1.5 Å along the helix axis and a rotation of 100 degrees, which gives 3.6 amino acid residues per turn of helix. Thus, amino acids spaced three and four apart in the sequence are spatially quite close to one another in an α helix. In contrast, amino acids spaced two apart in the sequence are situated on opposite sides of the helix and so are unlikely to make contact. The *pitch* of the α helix is the length of one complete turn along the helix axis and is equal to the product of the rise (1.5 Å) and the number of residues per turn (3.6), or 5.4 Å. The *screw*

Screw sense

Describes the direction in which a helical structure rotates with respect to its axis. If, viewed down the axis of a helix, the chain turns in a clockwise direction, it has a right-handed screw sense. If the turning is counterclockwise, the screw sense is left-handed.

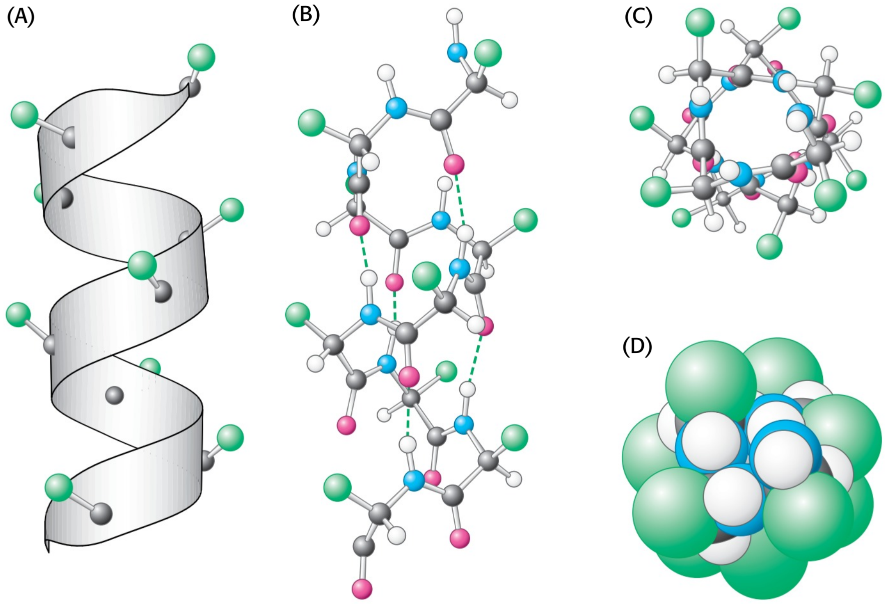


FIGURE 2.24 Structure of the α helix. (A) A ribbon depiction shows the α-carbon atoms and side chains (green). (B) A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. (C) An end view shows the coiled backbone as the inside of the helix and the side chains (green) projecting outward. (D) A space-filling view of part C shows the tightly packed interior core of the helix.

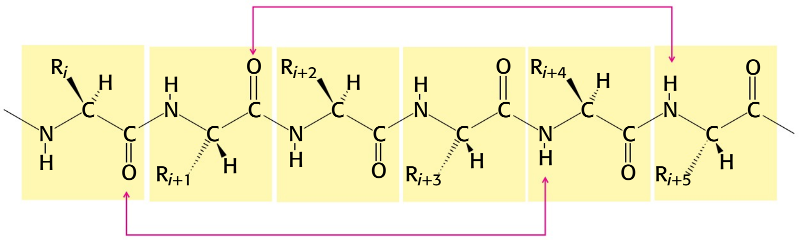


FIGURE 2.25 Hydrogen-bonding scheme for an α helix. In the α helix, the CO group of residue *i* forms a hydrogen bond with the NH group of residue *i* + 4.

sense of an α helix can be right-handed (clockwise) or left-handed (counterclockwise). The Ramachandran plot reveals that both the right-handed and the left-handed helices are among allowed conformations (Figure 2.26). However, right-handed helices are energetically more favorable because there is less steric clash between the side chains and the backbone. *Essentially all α helices found in proteins are right-handed.* In schematic representations of proteins, α helices are depicted as twisted ribbons or rods (Figure 2.27).

Not all amino acids can be readily accommodated in an α helix. Branching at the β-carbon atom, as in valine, threonine, and isoleucine, tends to destabilize α helices because of steric clashes. Serine, aspartate, and asparagine also tend to disrupt α helices because their side chains contain hydrogen-bond donors or acceptors in close proximity to the main chain, where they compete for main-chain NH and CO groups. Proline also is a helix breaker because it lacks an NH group and because its ring structure prevents it from assuming the ϕ value to fit into an α helix.

The α-helical content of proteins ranges widely, from none to almost 100%. For example, about 75% of the residues in ferritin, a protein that helps store iron, are in α helices (Figure 2.28). Indeed, about 25% of all soluble proteins are composed of α helices connected by loops and turns of the polypeptide chain. Single α helices are usually less than 45 Å long. Many proteins that span biological membranes also contain α helices.

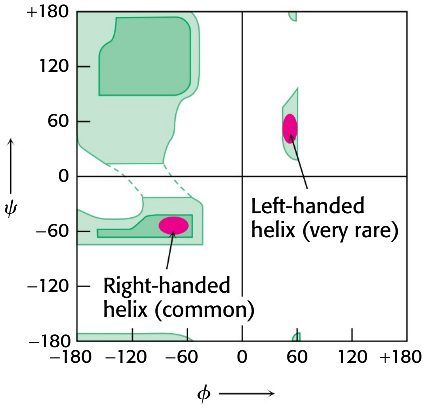


FIGURE 2.26 Ramachandran plot for helices. Both right- and left-handed helices lie in regions of allowed conformations in the Ramachandran plot. However, essentially all α helices in proteins are right-handed.

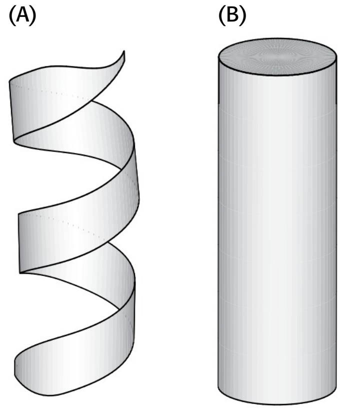


FIGURE 2.27 Schematic views of α helices. (A) A ribbon depiction. (B) A cylindrical depiction.

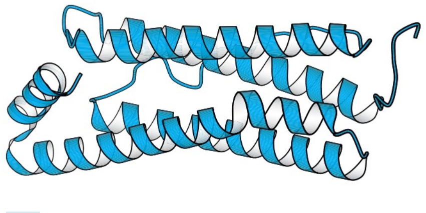


FIGURE 2.28 A largely α-helical protein. Ferritin, an iron-storage protein, is built from a bundle of α helices. [Drawn from 1AEW.pdb.]

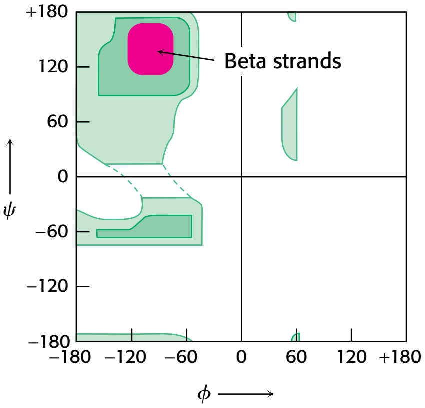


FIGURE 2.29 Ramachandran plot for β strands. The red area shows the sterically allowed conformations of extended, β-strand-like structures.

Beta sheets are stabilized by hydrogen bonding between polypeptide strands

Pauling and Corey proposed another periodic structural motif, which they named the *β pleated sheet* (β because it was the second structure that they elucidated, the α helix having been the first). The β pleated sheet (or, more simply, the β sheet) differs markedly from the rodlike α helix. It is composed of two or more polypeptide chains called *β strands*. A β strand is almost fully extended rather than being tightly coiled as in the α helix. A range of extended structures are sterically allowed (Figure 2.29).

The distance between adjacent amino acids along a β strand is approximately 3.5 Å, in contrast with a distance of 1.5 Å along an α helix. The side chains of adjacent amino acids point in opposite directions (Figure 2.30).

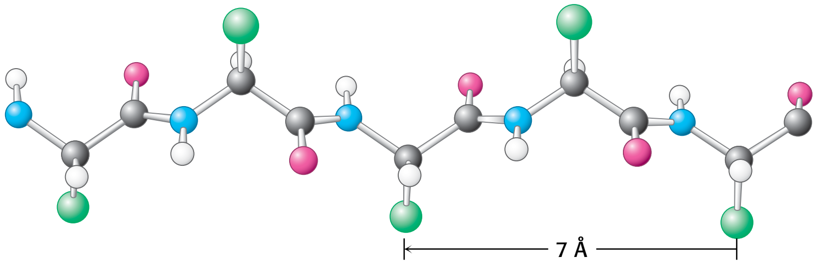


FIGURE 2.30 Structure of a β strand. The side chains (green) are alternately above and below the plane of the strand.

A β sheet is formed by linking two or more β strands lying next to one another through hydrogen bonds. Adjacent strands in a β sheet can run in opposite directions (antiparallel β sheet) or in the same direction (parallel β sheet). In the antiparallel arrangement, the NH group and the CO group of each amino acid are respectively hydrogen bonded to the CO group and the NH group of a partner on the adjacent chain (Figure 2.31). In the parallel arrangement, the hydrogen-bonding scheme is slightly more complicated. For each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on the adjacent strand, whereas the CO group is hydrogen bonded to the NH group on the amino acid two residues farther along the chain (Figure 2.32). Many strands, typically 4 or 5 but as many as 10 or more, can come together in β sheets. Such β sheets can be purely antiparallel, purely parallel, or mixed (Figure 2.33).

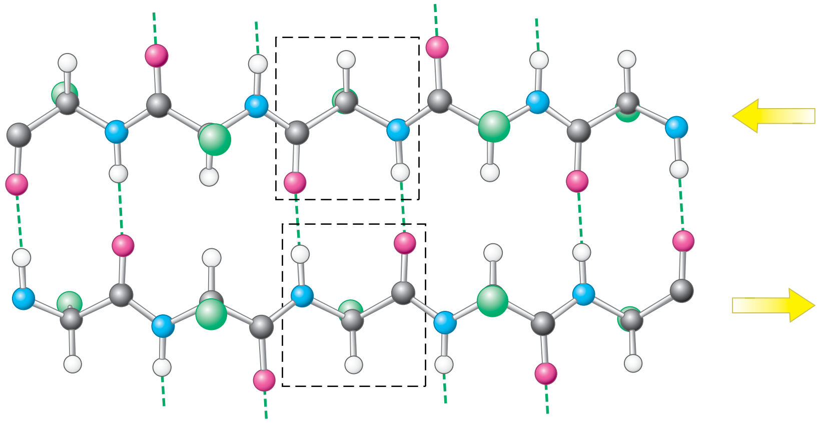


FIGURE 2.31 An antiparallel β sheet. Adjacent β strands run in opposite directions, as indicated by the arrows. Hydrogen bonds between NH and CO groups connect each amino acid to a single amino acid on an adjacent strand, stabilizing the structure.

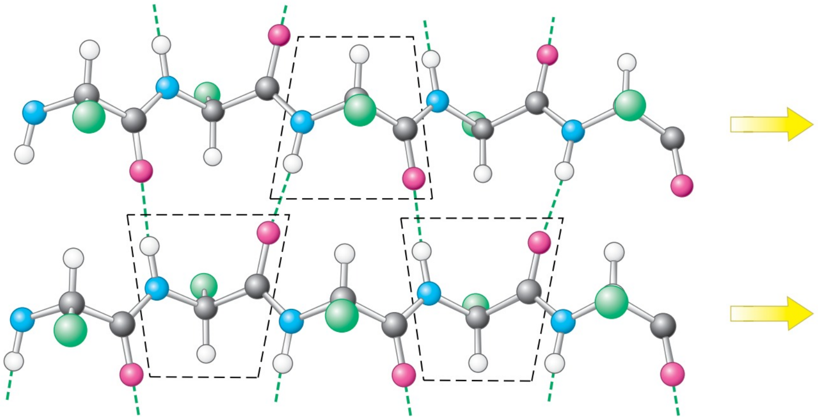


FIGURE 2.32 A parallel β sheet. Adjacent β strands run in the same direction, as indicated by the arrows. Hydrogen bonds connect each amino acid on one strand with two different amino acids on the adjacent strand.

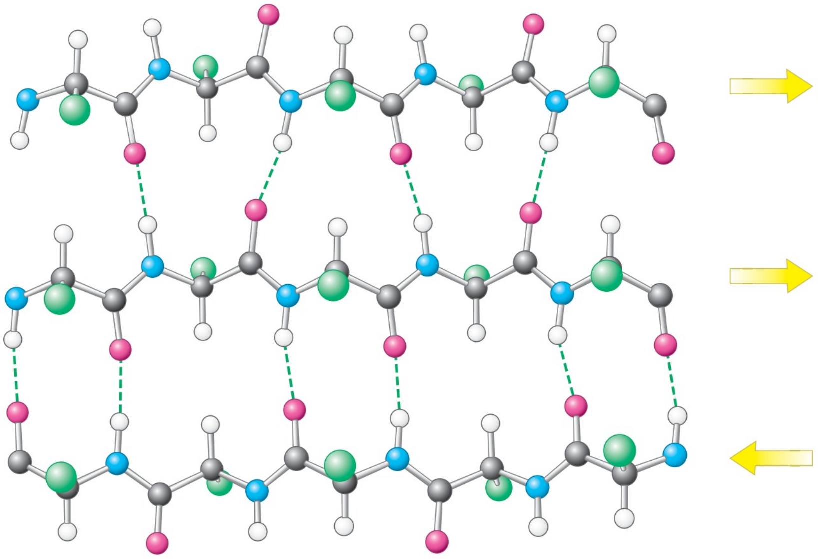


FIGURE 2.33 Structure of a mixed β sheet. The arrows indicate directionality of each strand.

In schematic representations, β strands are usually depicted by broad arrows pointing in the direction of the carboxyl-terminal end to indicate the type of β sheet formed—parallel or antiparallel. More structurally diverse than α helices, β sheets can be almost flat but most adopt a somewhat twisted shape (Figure 2.34). The β sheet is an important structural element in many proteins. For example, fatty acid-binding proteins, important for lipid metabolism, are built almost entirely from β sheets (Figure 2.35).

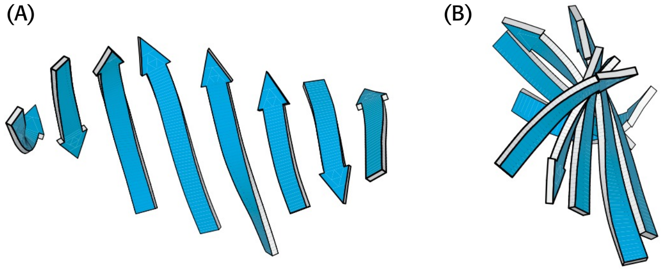


FIGURE 2.34 A schematic twisted β sheet. (A) A schematic model. (B) The schematic view rotated by 90 degrees to illustrate the twist more clearly.

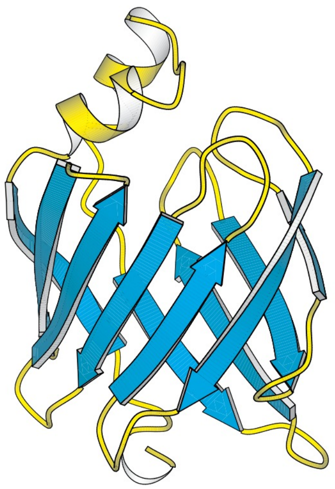


FIGURE 2.35 A protein rich in β sheets. The structure of a fatty acid-binding protein. [Drawn from 1FTP.pdb.]

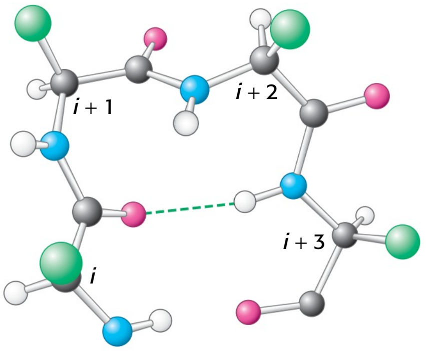


FIGURE 2.36 Structure of a reverse turn. The CO group of residue i of the polypeptide chain is hydrogen bonded to the NH group of residue $i + 3$ to stabilize the turn.

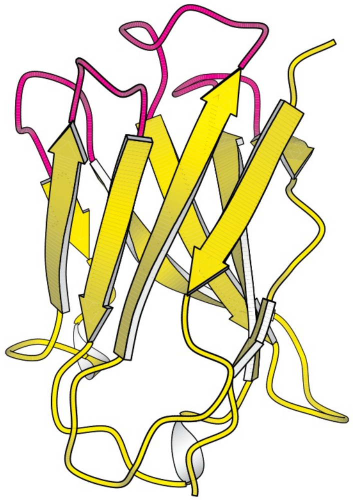


FIGURE 2.37 Loops on a protein surface. A part of an antibody molecule has surface loops (shown in red) that mediate interactions with other molecules. [Drawn from 7FAB.pdb.]

Polypeptide chains can change direction by making reverse turns and loops

Most proteins have compact, globular shapes owing to reversals in the direction of their polypeptide chains. Many of these reversals are accomplished by a common structural element called the *reverse turn* (also known as the *β turn* or *hairpin turn*), illustrated in Figure 2.36. In many reverse turns, the CO group of residue i of a polypeptide is hydrogen bonded to the NH group of residue $i + 3$. This interaction stabilizes abrupt changes in direction of the polypeptide chain. In other cases, more-elaborate structures are responsible for chain reversals. These structures are called *loops* or sometimes *Ω loops* (omega loops) to suggest their overall shape. Unlike α helices and β strands, loops do not have regular, periodic structures. Nonetheless, loop structures are often rigid and well defined (Figure 2.37). Turns and loops invariably lie on the surfaces of proteins and thus often participate in interactions between proteins and other molecules.

Fibrous proteins provide structural support for cells and tissues

Special types of helices are present in the two proteins α-keratin and collagen. These proteins form long fibers that serve a structural role.

α-Keratin, which is an essential component of wool, hair, and skin, consists of two right-handed α helices intertwined to form a type of left-handed superhelix called an *α-helical coiled coil*. α-Keratin is a member of a superfamily of proteins referred to as *coiled-coil proteins* (Figure 2.38). In these proteins, two or more α helices can entwine to form a very stable structure, which can have a length of 1000 Å (100 nm, or 0.1 μm) or more. There are approximately 60 members of this family in humans, including intermediate filaments, proteins that contribute to the cell cytoskeleton (internal scaffolding in a cell), and the muscle proteins myosin and tropomyosin (Section 35.2). Members of this family are characterized by a central region of 300 amino acids that contains imperfect repeats of a sequence of seven amino acids called a *heptad repeat*.

The two helices in α-keratin associate with each other by weak interactions such as van der Waals forces and ionic interactions. The left-handed supercoil alters the two right-handed α helices such that there are 3.5 residues per turn instead of 3.6. Thus, the pattern of side-chain interactions can be repeated every seven residues, forming the heptad repeats. Two helices with such repeats are able to interact with one another if the repeats are complementary (Figure 2.39). For example, the repeating residues may be hydrophobic, allowing van der Waals interactions, or have opposite charge, allowing ionic interactions. In addition, the two helices may be linked by disulfide bonds formed by neighboring cysteine residues. The bonding of the helices accounts for the physical properties of wool, an example of an α-keratin. Wool is extensible and can be stretched to nearly twice its length because the α helices stretch, breaking

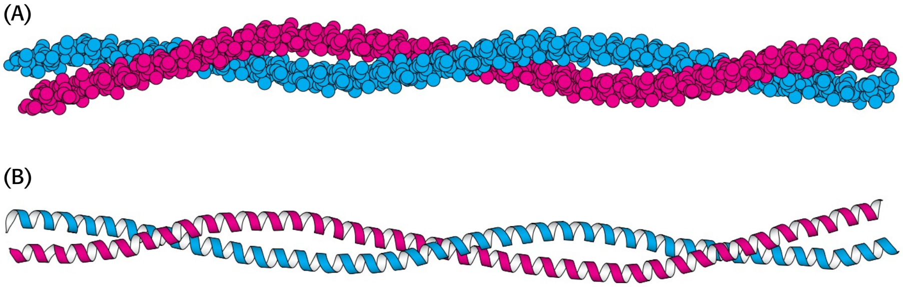


FIGURE 2.38 An α-helical coiled coil. (A) Space-filling model. (B) Ribbon diagram. The two helices wind around one another to form a superhelix. Such structures are found in many proteins, including keratin in hair, quills, claws, and horns. [Drawn from 1C1G.pdb.]

the weak interactions between neighboring helices. However, the covalent disulfide bonds resist breakage and return the fiber to its original state once the stretching force is released. The number of disulfide bond cross-links further defines the fiber's properties. Hair and wool, having fewer cross-links, are flexible. Horns, claws, and hooves, having more cross-links, are much harder.

A different type of helix is present in collagen, the most abundant protein of mammals. Collagen is the main fibrous component of skin, bone, tendon, cartilage, and teeth. This extracellular protein is a rod-shaped molecule, about 3000 Å long and only 15 Å in diameter. It contains three helical polypeptide chains, each nearly 1000 residues long. Glycine appears at every third residue in the amino acid sequence, and the sequence glycine-proline-hydroxyproline recurs frequently (Figure 2.40). Hydroxyproline is a derivative of proline that has a hydroxyl group in place of one of the hydrogen atoms on the pyrrolidine ring.

The collagen helix has properties different from those of the α helix. Hydrogen bonds within a strand are absent. Instead, *the helix is stabilized by steric repulsion of the pyrrolidine rings of the proline and hydroxyproline residues* (Figure 2.41). The pyrrolidine rings keep out of each other's way when the polypeptide chain assumes its helical form, which has about three residues per turn. Three strands wind around one another to form a *superhelical cable* that is stabilized by hydrogen bonds between strands. The hydrogen bonds form between the peptide NH groups of glycine residues and the CO groups of residues on the other chains. The hydroxyl groups of hydroxyproline residues also participate in hydrogen bonding.

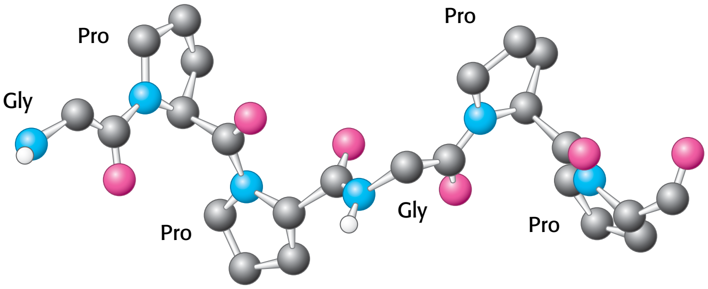


FIGURE 2.41 Conformation of a single strand of a collagen triple helix.

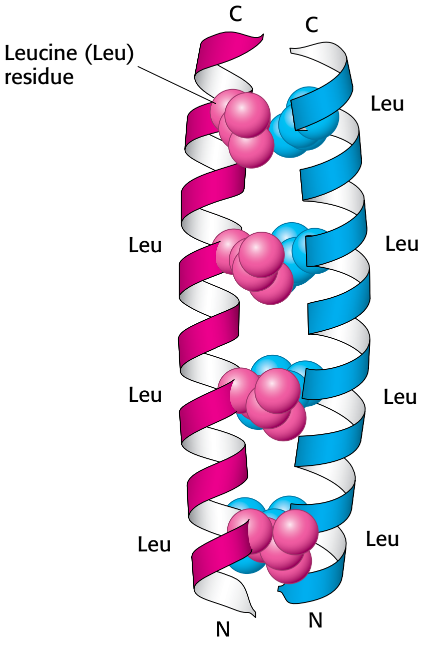


FIGURE 2.39 Heptad repeats in a coiled-coil protein. Every seventh residue in each helix is leucine. The two helices are held together by van der Waals interactions primarily between the leucine residues. [Drawn from 2ZTA.pdb.]

```
13
-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-
22
-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Ala-Hyp-
31
-Gly-Pro-Gln-Gly-Phe-Gln-Gly-Pro-Hyp-
40
-Gly-Glu-Hyp-Gly-Glu-Hyp-Gly-Ala-Ser-
49
-Gly-Pro-Met-Gly-Pro-Arg-Gly-Pro-Hyp-
58
-Gly-Pro-Hyp-Gly-Lys-Asn-Gly-Asp-Asp-
```

FIGURE 2.40 Amino acid sequence of a part of a collagen chain. Every third residue is a glycine. Proline and hydroxyproline (Hyp) also are abundant.

The inside of the triple-stranded helical cable is very crowded and accounts for the requirement that glycine be present at every third position on each strand (Figure 2.42A). *The only residue that can fit in an interior position is glycine*. The amino acid residue on either side of glycine is located on the outside of the cable, where there is room for the bulky rings of proline and hydroxyproline residues (Figure 2.42B).

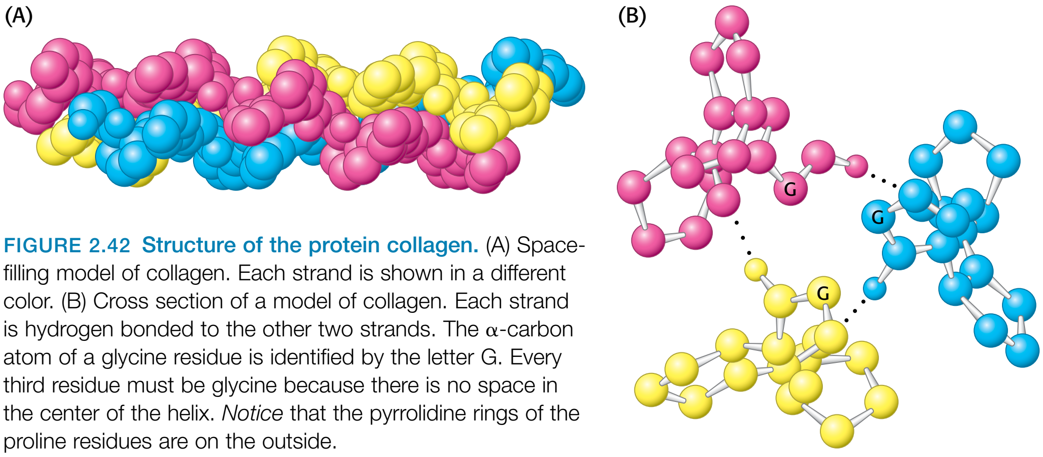


FIGURE 2.42 Structure of the protein collagen. (A) Space-filling model of collagen. Each strand is shown in a different color. (B) Cross section of a model of collagen. Each strand is hydrogen bonded to the other two strands. The α-carbon atom of a glycine residue is identified by the letter G. Every third residue must be glycine because there is no space in the center of the helix. *Notice* that the pyrrolidine rings of the proline residues are on the outside.

The importance of the positioning of glycine inside the triple helix is illustrated in the disorder osteogenesis imperfecta, also known as brittle bone disease. In this condition, which can vary from mild to very severe, other amino acids replace the internal glycine residue. This replacement leads to a delayed and improper folding of collagen. The most serious symptom is severe bone fragility. Defective collagen in the eyes causes the whites of the eyes to have a blue tint (blue sclera).

2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores

Let us now examine how amino acids are grouped together in a complete protein. X-ray crystallographic and nuclear magnetic resonance (NMR) studies (Section 3.5) have revealed the detailed three-dimensional structures of thousands of proteins. We begin here with an examination of *myoglobin,* the first protein to be seen in atomic detail.

Myoglobin, the oxygen storage protein in muscle, is a single polypeptide chain of 153 amino acids (Chapter 7). The capacity of myoglobin to bind oxygen depends on the presence of *heme,* a nonpolypeptide *prosthetic (helper) group* consisting of protoporphyrin IX and a central iron atom. *Myoglobin is an extremely compact molecule.* Its overall dimensions are 45 × 35 × 25 Å, an order of magnitude less than if it were fully stretched out (Figure 2.43). About 70% of the main chain is folded into eight α helices, and much of the rest of the chain forms turns and loops between helices.

The folding of the main chain of myoglobin, like that of most other proteins, is complex and devoid of symmetry. The overall course of the polypeptide chain of a protein is referred to as its *tertiary structure.* A unifying principle emerges from the distribution of side chains. Strikingly, *the interior consists almost entirely of nonpolar residues* such as leucine, valine, methionine, and phenylalanine (Figure 2.44). Charged residues such as aspartate, glutamate, lysine, and arginine are absent from the inside of myoglobin. The only polar residues inside are two histidine residues, which play critical roles in binding iron and oxygen. The outside of myoglobin, on the

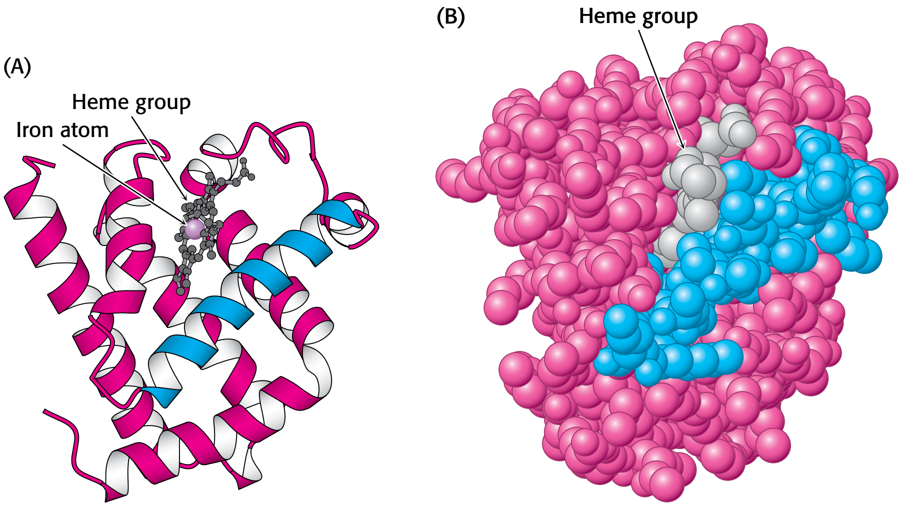


FIGURE 2.43 Three-dimensional structure of myoglobin. (A) A ribbon diagram shows that the protein consists largely of α helices. (B) A space-filling model in the same orientation shows how tightly packed the folded protein is. *Notice* that the heme group is nestled into a crevice in the compact protein with only an edge exposed. One helix is blue to allow comparison of the two structural depictions. [Drawn from 1A6N.pdb.]

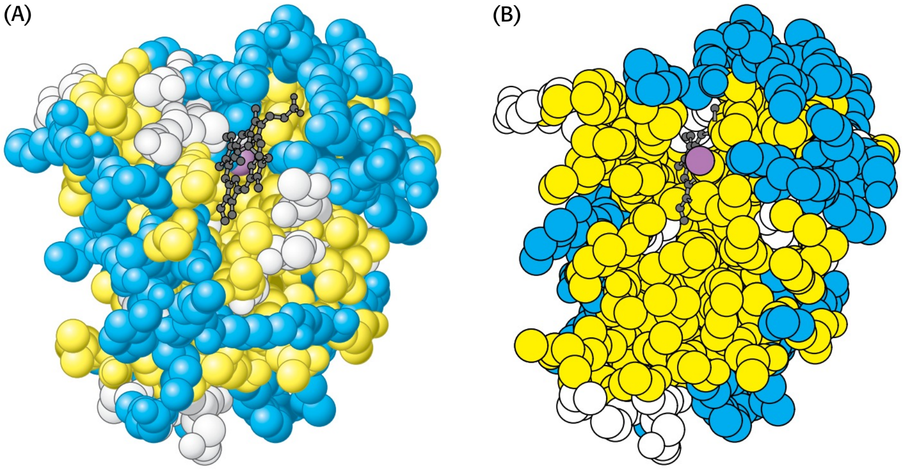


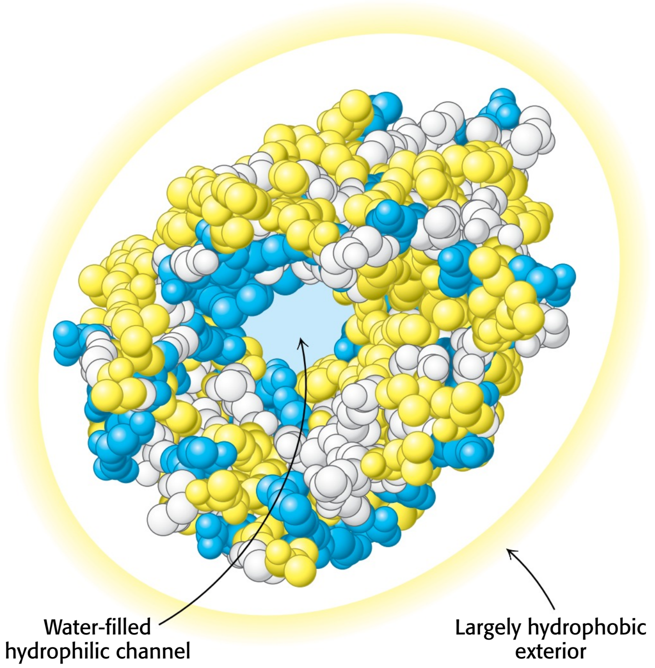
FIGURE 2.44 Distribution of amino acids in myoglobin. (A) A space-filling model of myoglobin with hydrophobic amino acids shown in yellow, charged amino acids shown in blue, and others shown in white. *Notice* that the surface of the molecule has many charged amino acids, as well as some hydrophobic amino acids. (B) In this cross-sectional view, *notice* that mostly hydrophobic amino acids are found on the inside of the structure, whereas the charged amino acids are found on the protein surface. [Drawn from 1MBD.pdb.]

other hand, consists of both polar and nonpolar residues. The space-filling model shows that there is very little empty space inside.

This contrasting distribution of polar and nonpolar residues reveals a key facet of protein architecture. In an aqueous environment, protein folding is driven by the strong tendency of hydrophobic residues to be excluded from water. Recall that a system is more thermodynamically stable when hydrophobic groups are clustered rather than extended into the aqueous surroundings (p. 9). *The polypeptide chain therefore folds so that its hydrophobic side chains are buried and its polar, charged chains are on the surface*. Many α helices and β strands are amphipathic; that is, the α helix or β strand has a hydrophobic face, which points into the protein interior, and a more polar face, which points into solution. The fate of the main chain accompanying the hydrophobic side chains is important, too. An unpaired peptide NH or CO group markedly prefers water to a nonpolar milieu. The secret of burying a segment of main chain in a hydrophobic environment is to pair all the NH and CO groups by hydrogen bonding. This pairing is neatly accomplished in an α helix or β sheet. Van der Waals interactions between tightly packed hydrocarbon side chains also contribute to the stability of proteins. We can now understand why the set of 20 amino acids contains several that differ subtly in size and shape. They provide a palette from which to choose to fill the interior of a protein neatly and thereby maximize van der Waals interactions, which require intimate contact.

Some proteins that span biological membranes are "the exceptions that prove the rule" because they have the reverse distribution of hydrophobic and hydrophilic amino acids. For example, consider porins, proteins found in the outer membranes of many bacteria (Figure 2.45). Membranes are built largely of hydrophobic alkane chains (Section 12.2). Thus, porins are covered on the outside largely with hydrophobic residues that interact with the neighboring alkane chains. In contrast, the center of the protein contains many charged and polar amino acids that surround a water-filled channel running through the middle of the protein. Thus, because porins function in hydrophobic environments, they are "inside out" relative to proteins that function in aqueous solution.

FIGURE 2.45 "Inside out" amino acid distribution in porin. The outside of porin (which contacts hydrophobic groups in membranes) is covered largely with hydrophobic residues, whereas the center includes a water-filled channel lined with charged and polar amino acids. [Drawn from 1PRN.pdb.]


Water-filled hydrophilic channel
Largely hydrophobic exterior

Certain combinations of secondary structure are present in many proteins and frequently exhibit similar functions. These combinations are called *motifs* or *supersecondary structures*. For example, an α helix separated from another α helix by a turn, called a *helix-turn-helix* unit, is found in many proteins that bind DNA (Figure 2.46).

Some polypeptide chains fold into two or more compact regions that may be connected by a flexible segment of polypeptide chain, rather like pearls on a string. These compact globular units, called *domains,* range in size from about 30 to 400 amino acid residues. For example, the extracellular part of CD4, a protein on the surface of certain cells of the immune system (Section 34.4), comprises four similar domains of approximately 100 amino acids each (Figure 2.47). Proteins may have domains in common even if their overall tertiary structures are different.

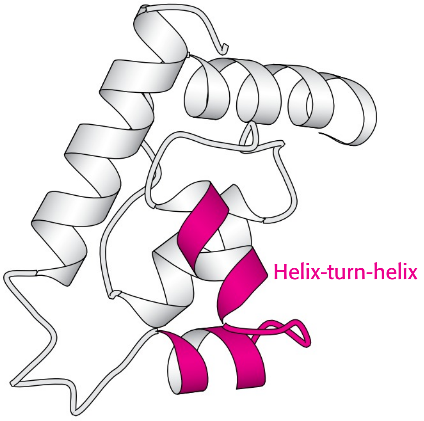


FIGURE 2.46 The helix-turn-helix motif, a supersecondary structural element. Helix-turn-helix motifs are found in many DNA-binding proteins. [Drawn from 1LMB.pdb.]

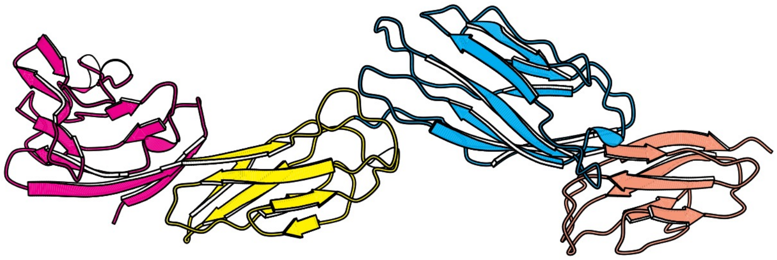


FIGURE 2.47 Protein domains. The cell-surface protein CD4 consists of four similar domains. [Drawn from 1WIO.pdb.]

2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures

Four levels of structure are frequently cited in discussions of protein architecture. So far, we have considered three. *Primary structure* is the amino acid sequence. *Secondary structure* refers to the spatial arrangement of amino acid residues that are nearby in the sequence. Some of these arrangements give rise to periodic structures. The α helix and β strand are elements of secondary structure. *Tertiary structure* refers to the spatial arrangement of amino acid residues that are far apart in the sequence and to the pattern of disulfide bonds. We now turn to proteins containing more than one polypeptide chain. Such proteins exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a *subunit*. *Quaternary structure* refers to the spatial arrangement of subunits and the nature of their interactions. The simplest sort of quaternary structure is a *dimer,* consisting of two identical subunits. This organization is present in the DNA-binding protein Cro, found in a bacterial virus called λ (Figure 2.48). More-complicated quaternary structures also are common. More than one type of subunit can be present, often in variable numbers. For example, human hemoglobin, the oxygen-carrying protein in blood, consists of two subunits of one type (designated α) and two subunits of another type (designated β), as illustrated in Figure 2.49. Thus, the hemoglobin molecule exists as an $\alpha_2\beta_2$ tetramer. Subtle changes in the arrangement of subunits within the hemoglobin molecule allow it to carry oxygen from the lungs to tissues with great efficiency (Chapter 7).

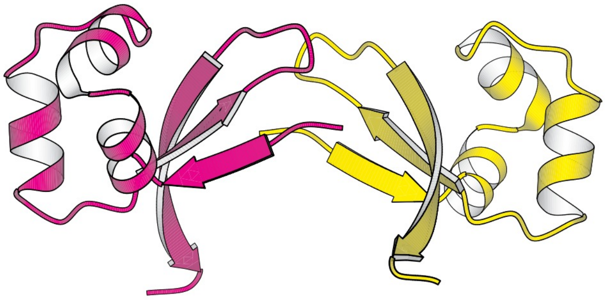


FIGURE 2.48 Quaternary structure. The Cro protein of bacteriophage λ is a dimer of identical subunits. [Drawn from 5CRO.pdb.]

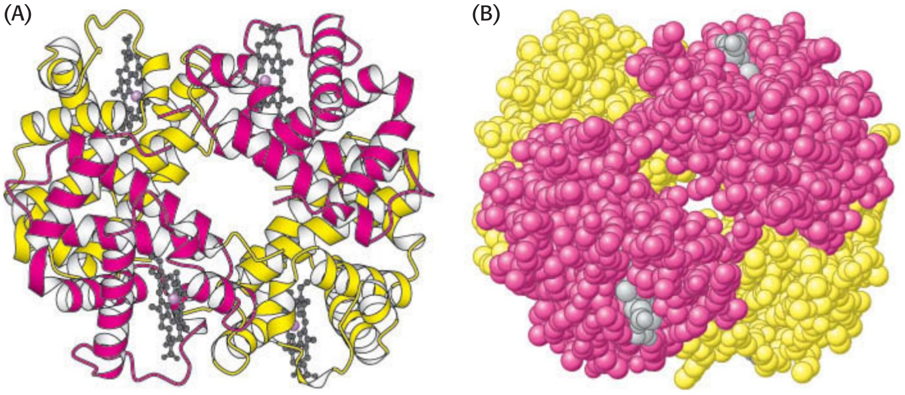


FIGURE 2.49 The $\alpha_2\beta_2$ tetramer of human hemoglobin. The structure of the two α subunits (red) is similar to but not identical with that of the two β subunits (yellow). The molecule contains four heme groups (gray with the iron atom shown in purple). (A) The ribbon diagram highlights the similarity of the subunits and shows that they are composed mainly of α helices. (B) The space-filling model illustrates how the heme groups occupy crevices in the protein. [Drawn from 1A3N.pdb.]

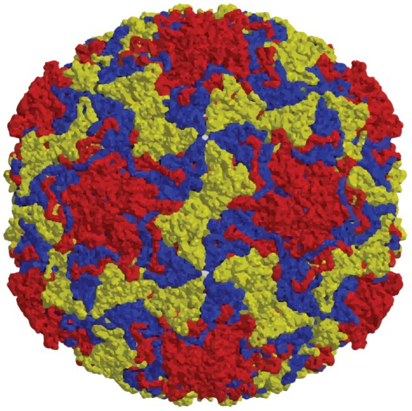


FIGURE 2.50 Complex quaternary structure. The coat of human rhinovirus, the cause of the common cold, comprises 60 copies of each of four subunits. The three most prominent subunits are shown as different colors.

Viruses make the most of a limited amount of genetic information by forming coats that use the same kind of subunit repetitively in a symmetric array. The coat of rhinovirus, the virus that causes the common cold, includes 60 copies of each of four subunits (Figure 2.50). The subunits come together to form a nearly spherical shell that encloses the viral genome.

2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

How is the elaborate three-dimensional structure of proteins attained? The classic work of Christian Anfinsen in the 1950s on the enzyme ribonuclease revealed the relation between the amino acid sequence of a protein and its conformation. Ribonuclease is a single polypeptide chain consisting of 124 amino acid residues cross-linked by four disulfide bonds (Figure 2.51). Anfinsen's plan was to destroy the three-dimensional structure of the enzyme and to then determine what conditions were required to restore the structure.

Agents such as *urea* or *guanidinium chloride* effectively disrupt a protein's noncovalent bonds. Although the mechanism of action of these agents is not fully understood, computer simulations suggest that they replace water as the molecule solvating the protein and are then able to disrupt the van der Waals interactions stabilizing the protein structure. The disulfide bonds can be cleaved reversibly by reducing them with a reagent such as *β-mercaptoethanol* (Figure 2.52). In the presence of a large excess of β-mercaptoethanol, the disulfides (cystines) are fully converted into sulfhydryls (cysteines).

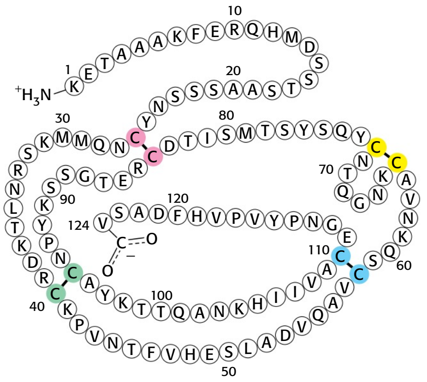
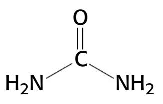
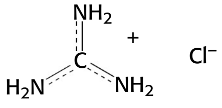

^+H_3N
1
10
20
30
40
50
60
70
80
90
100
110
120
124
C=O
O
−
KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV

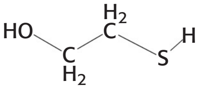
FIGURE 2.51 Amino acid sequence of bovine ribonuclease. The four disulfide bonds are shown in color. [After C. H. W. Hirs, S. Moore, and W. H. Stein, *J. Biol. Chem.* 235:633–647, 1960.]



Urea



Guanidinium chloride



β-Mercaptoethanol

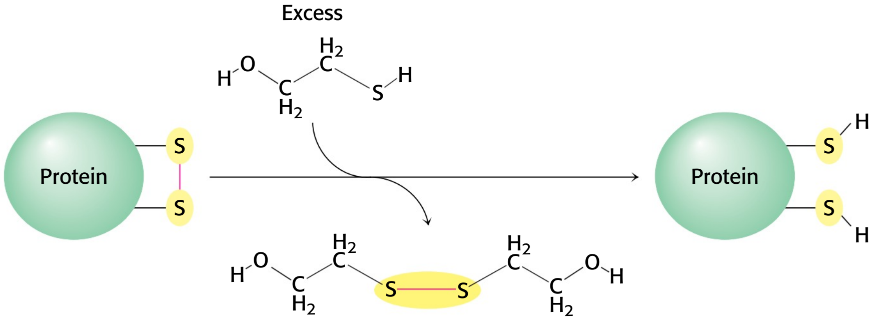
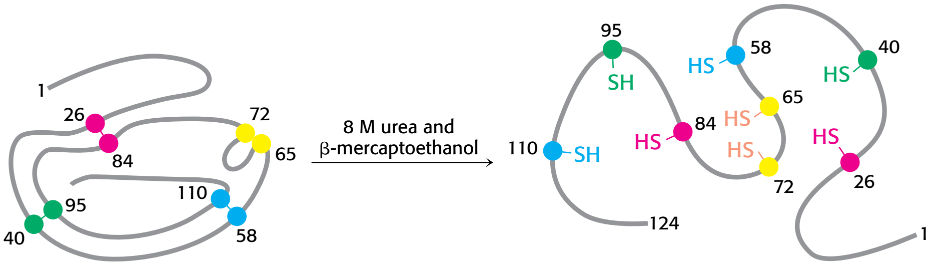


FIGURE 2.52 Role of β-mercaptoethanol in reducing disulfide bonds. Note that, as the disulfides are reduced, the β-mercaptoethanol is oxidized and forms dimers.

FIGURE 2.53 Reduction and denaturation of ribonuclease.



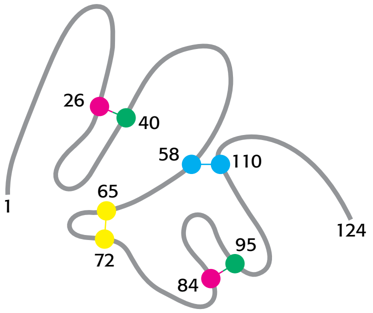
Native ribonuclease **Denatured reduced ribonuclease**

Most polypeptide chains devoid of cross-links assume a *random-coil conformation* in 8 M urea or 6 M guanidinium chloride. When ribonuclease was treated with β-mercaptoethanol in 8 M urea, the product was a fully reduced, randomly coiled polypeptide chain *devoid of enzymatic activity*. When a protein is converted into a randomly coiled peptide without its normal activity, it is said to be *denatured* (Figure 2.53).

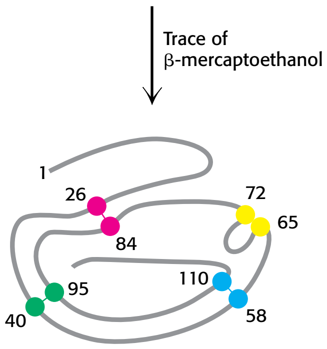
Anfinsen then made the critical observation that the denatured ribonuclease, freed of urea and β-mercaptoethanol by dialysis (Section 3.1), slowly regained enzymatic activity. He perceived the significance of this chance finding: the sulfhydryl groups of the denatured enzyme became oxidized by air, and the enzyme spontaneously refolded into a catalytically active form. Detailed studies then showed that nearly all the original enzymatic activity was regained if the sulfhydryl groups were oxidized under suitable conditions. All the measured physical and chemical properties of the refolded enzyme were virtually identical with those of the native enzyme. These experiments showed that *the information needed to specify the catalytically active structure of ribonuclease is contained in its amino acid sequence*. Subsequent studies have established the generality of this central principle of biochemistry: *sequence specifies conformation*. The dependence of conformation on sequence is especially significant because of the intimate connection between conformation and function.

A quite different result was obtained when reduced ribonuclease was reoxidized while it was still in 8 M urea and the preparation was then dialyzed to remove the urea. Ribonuclease reoxidized in this way had only 1% of the enzymatic activity of the native protein. Why were the outcomes so different when reduced ribonuclease was reoxidized in the presence and absence of urea? The reason is that the wrong disulfides formed pairs in urea. There are 105 different ways of pairing eight cysteine molecules to form four disulfides; only one of these combinations is enzymatically active. The 104 wrong pairings have been picturesquely termed "scrambled" ribonuclease. Anfinsen found that scrambled ribonuclease spontaneously converted into fully active, native ribonuclease when trace amounts of β-mercaptoethanol were added to an aqueous solution of the protein (Figure 2.54). The added β-mercaptoethanol catalyzed the rearrangement of disulfide pairings until the native structure was regained in about 10 hours. *This process was driven by the decrease in free energy as the scrambled conformations were converted into the stable, native conformation of the enzyme.* The native disulfide pairings of ribonuclease thus contribute to the stabilization of the thermodynamically preferred structure.

Similar refolding experiments have been performed on many other proteins. In many cases, the native structure can be generated under suitable conditions. For other proteins, however, refolding does not proceed efficiently. In these cases, the unfolded protein molecules usually become tangled up with one another to form aggregates. Inside cells, proteins called *chaperones* block such undesirable interactions. Additionally, it is now



Scrambled ribonuclease



Native ribonuclease

FIGURE 2.54 Reestablishing correct disulfide pairing. Native ribonuclease can be re-formed from scrambled ribonuclease in the presence of a trace of β-mercaptoethanol.

evident that some proteins do not assume a defined structure until they interact with molecular partners, as we will see shortly.

Amino acids have different propensities for forming α helices, β sheets, and turns

How does the amino acid sequence of a protein specify its three-dimensional structure? How does an unfolded polypeptide chain acquire the form of the native protein? These fundamental questions in biochemistry can be approached by first asking a simpler one: What determines whether a particular sequence in a protein forms an α helix, a β strand, or a turn? One source of insight is to examine the frequency of occurrence of particular amino acid residues in these secondary structures (Table 2.3). Residues such as alanine, glutamate, and leucine tend to be present in α helices, whereas valine and isoleucine tend to be present in β strands. Glycine, asparagine, and proline are more commonly observed in turns.

Studies of proteins and synthetic peptides have revealed some reasons for these preferences. Branching at the β-carbon atom, as in valine, threonine, and isoleucine, tends to destabilize α helices because of steric clashes. These residues are readily accommodated in β strands, where their side chains project out of the plane containing the main chain. Serine and asparagine tend to disrupt α helices because their side chains contain hydrogen-bond donors or acceptors in close proximity to the main chain, where they compete for main-chain NH and CO groups. Proline tends to disrupt both α helices and β strands because it lacks an NH group and because its ring structure restricts its ϕ value to near 60 degrees. Glycine readily fits into all structures, but its conformational flexibility renders it well-suited to reverse turns.

Can we predict the secondary structure of a protein by using this knowledge of the conformational preferences of amino acid residues?

TABLE 2.3 Relative frequencies of amino acid residues in secondary structures

Amino acid	α helix	β sheet	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.96
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (middle group), or turns (bottom group).

Source: T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.

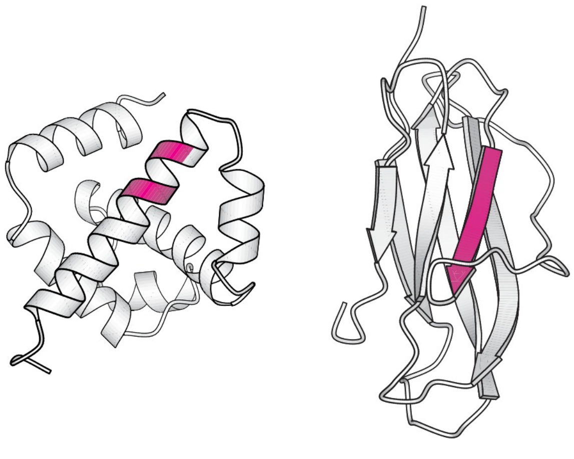


FIGURE 2.55 Alternative conformations of a peptide sequence. Many sequences can adopt alternative conformations in different proteins. Here the sequence VDLLKN shown in red assumes an α helix in one protein context (left) and a β strand in another (right). [Drawn from (left) 3WRP.pdb and (right) 2HLA.pdb.]

Accurate predictions of secondary structure adopted by even a short stretch of residues have proved to be difficult. What stands in the way of more-accurate prediction? Note that the conformational preferences of amino acid residues are not tipped all the way to one structure (Table 2.3). For example, glutamate, one of the strongest helix formers, prefers α helix to β strand by only a factor of three. The preference ratios of most other residues are smaller. Indeed, some penta- and hexapeptide sequences have been found to adopt one structure in one protein and an entirely different structure in another (Figure 2.55). Hence, some amino acid sequences do not uniquely determine secondary structure. Tertiary interactions—interactions between residues that are far apart in the sequence—may be decisive in specifying the secondary structure of some segments. Context is often crucial: the conformation of a protein has evolved to work in a particular environment. Nevertheless, substantial improvements in secondary structure prediction have been achieved by using families of related sequences, each of which adopts the same structure.

Protein folding is a highly cooperative process

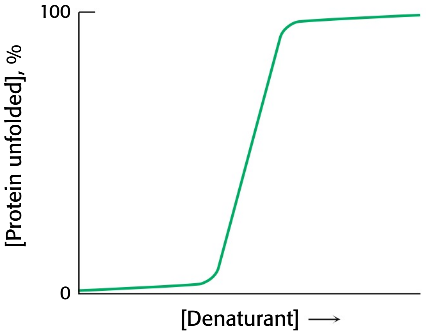


FIGURE 2.56 Transition from folded to unfolded state. Most proteins show a sharp transition from the folded to the unfolded form on treatment with increasing concentrations of denaturants.

Proteins can be denatured by any treatment that disrupts the weak bonds stabilizing tertiary structure, such as heating, or by chemical denaturants such as urea or guanidinium chloride. For many proteins, a comparison of the degree of unfolding as the concentration of denaturant increases reveals a sharp transition from the folded, or native, form to the unfolded, or denatured form, suggesting that only these two conformational states are present to any significant extent (Figure 2.56). A similar sharp transition is observed if denaturants are removed from unfolded proteins, allowing the proteins to fold.

The sharp transition seen in Figure 2.56 suggests that protein folding and unfolding is an *"all or none" process* that results from a *cooperative transition*. For example, suppose that a protein is placed in conditions under which some part of the protein structure is thermodynamically unstable. As this part of the folded structure is disrupted, the interactions between it and the remainder of the protein will be lost. The loss of these interactions, in turn, will destabilize the remainder of the structure. Thus, conditions that lead to the disruption of any part of a protein structure are likely to unravel the protein completely. The structural properties of proteins provide a clear rationale for the cooperative transition.

The consequences of cooperative folding can be illustrated by considering the contents of a protein solution under conditions corresponding to the middle of the transition between the folded and the unfolded forms. Under these conditions, the protein is "half folded." Yet the solution will appear to have no partly folded molecules but, instead, look as if it is a 50/50 mixture of fully folded and fully unfolded molecules (Figure 2.57). Although the protein may appear to behave as if it exists in only two states, this simple two-state existence is an impossibility at a molecular level. Even simple reactions go through reaction intermediates, and so a complex molecule such as a protein cannot simply switch from a completely unfolded state to the native state in one step. Unstable, transient

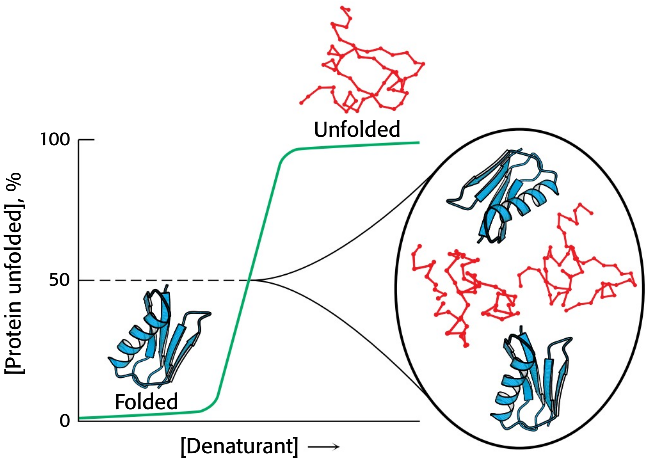


FIGURE 2.57 Components of a partly denatured protein solution. In a half-unfolded protein solution, half the molecules are fully folded and half are fully unfolded.

intermediate structures must exist between the native and denatured state (p. 54). Determining the nature of these intermediate structures is an area of intense biochemical research.

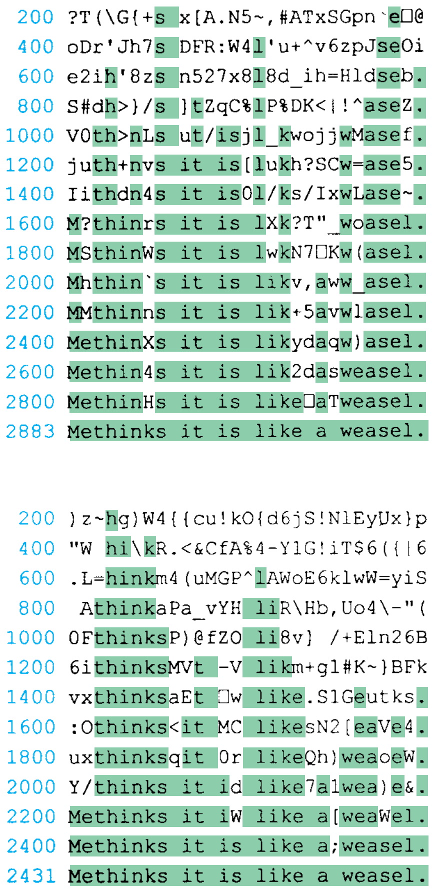
Proteins fold by progressive stabilization of intermediates rather than by random search

How does a protein make the transition from an unfolded structure to a unique conformation in the native form? One possibility a priori would be that all possible conformations are sampled to find the energetically most favorable one. How long would such a random search take? Consider a small protein with 100 residues. Cyrus Levinthal calculated that, if each residue can assume three different conformations, the total number of structures would be 3^{100}, which is equal to 5×10^{47}. If it takes 10^{-13} s to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or 1.6×10^{27} years. In reality, small proteins can fold in less than a second. Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations. The enormous difference between calculated and actual folding times is called *Levinthal's paradox*. This paradox clearly reveals that proteins do not fold by trying every possible conformation; instead, they must follow at least a partly defined folding pathway consisting of intermediates between the fully denatured protein and its native structure.

The way out of this paradox is to recognize the power of *cumulative selection*. Richard Dawkins, in *The Blind Watchmaker,* asked how long it would take a monkey poking randomly at a typewriter to reproduce Hamlet's remark to Polonius, "Methinks it is like a weasel" (Figure 2.58). An astronomically large number of keystrokes, on the order of 10^{40}, would be required. However, suppose that we preserved each correct character and allowed the monkey to retype only the wrong ones. In this case, only a few thousand keystrokes, on average, would be needed. The crucial difference between these cases is that the first employs a completely random search, whereas, in the second, *partly correct intermediates are retained.*

The essence of protein folding is the tendency to retain partly correct intermediates. However, the protein-folding problem is much more difficult than the one presented to our simian Shakespeare. First, the criterion of correctness is not a residue-by-residue scrutiny of conformation by an omniscient observer but rather the total free energy of the transient species. Second, proteins are only marginally stable. The free-energy difference between the folded and the unfolded states of a typical 100-residue protein is 42 kJ mol^{-1} (10 kcal mol^{-1}), and thus each residue contributes on average only 0.42 kJ mol^{-1} (0.1 kcal mol^{-1}) of energy to maintain the folded state. This amount is less than the amount of thermal energy, which is 2.5 kJ mol^{-1} (0.6 kcal mol^{-1}) at room temperature. This meager stabilization energy means that correct intermediates, especially those formed early in folding, can be lost. The analogy is that the monkey would be somewhat free to undo its correct keystrokes. Nonetheless, the interactions that lead to cooperative folding can stabilize intermediates as structure builds up. Thus, local regions that have significant structural preference, though not necessarily stable on their own, will tend to adopt their favored structures and, as they form, can interact with one other, leading to increasing stabilization. This conceptual framework is often referred to as the *nucleation-condensation model*.

A simulation of the folding of a protein, based on the nucleation-condensation model, is shown in Figure 2.59. This model suggests that



```
 200 ?T(\G{+s x[A.N5~,#ATxSGpn`e□@
 400 oDr'Jh7s DFR:W4l'u+^v6zpJseOi
 600 e2ih'8zs n527x8l8d_ih=Hldseb.
 800 S#dh>}/s }tZqC%lP%DK<¦!^aseZ.
1000 V0th>nLs ut/isjl_kwojjwMasef.
1200 juth+nvs it is[lukh?SCw=ase5.
1400 Iithdn4s it is0l/ks/IxwLase~.
1600 M?thinrs it is lXk?T"_woasel.
1800 MSthinWs it is lwkN7□Kw(asel.
2000 Mhthin`s it is likv,aww_asel.
2200 MMthinns it is lik+5avwlasel.
2400 MethinXs it is likydaqw)asel.
2600 Methin4s it is lik2dasweasel.
2800 MethinHs it is like□aTweasel.
2883 Methinks it is like a weasel.
```

```
 200 )z~hg)W4{{cu!kO{d6jS!NlEyUx}p
 400 "W hi\kR.<&CfA%4-Y1G!iT$6({|6
 600 .L=hinkm4(uMGP^lAWoE6klwW=yiS
 800  AthinkaPa_vYH liR\Hb,Uo4\-"(
1000 0FthinksP)@fZO li8v} /+Eln26B
1200 6ithinksMVt -V likm+gl#K~}BFk
1400 vxthinksaEt □w like.S1Geutks.
1600 :Othinks<it MC likesN2[eaVe4.
1800 uxthinksqit 0r likeQh)weaoeW.
2000 Y/thinks it id like7alwea)e&.
2200 Methinks it iW like a[weaWel.
2400 Methinks it is like a;weasel.
2431 Methinks it is like a weasel.
```

FIGURE 2.58 Typing-monkey analogy. A monkey randomly poking a typewriter could write a line from Shakespeare's *Hamlet,* provided that correct keystrokes were retained. In the two computer simulations shown, the cumulative number of keystrokes is given at the left of each line.

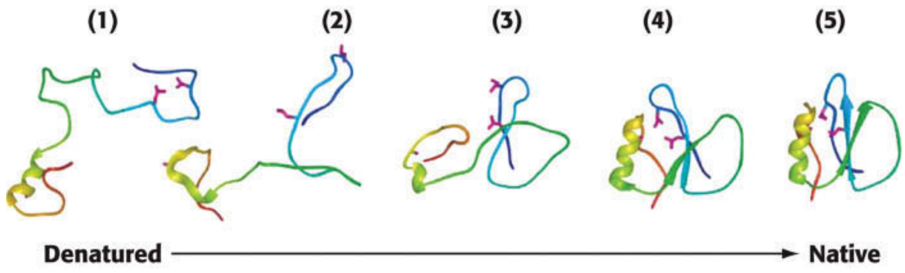


FIGURE 2.59 Proposed folding pathway of chymotrypsin inhibitor. Local regions with sufficient structural preference tend to adopt their favored structures initially (1). These structures come together to form a nucleus with a nativelike, but still mobile, structure (4). This structure then fully condenses to form the native, more rigid structure (5). [From A. R. Fersht and V. Daggett. *Cell* 108:573–582, 2002; with permission from Elsevier.]

certain pathways may be preferred. Although Figure 2.59 suggests a discrete pathway, each of the intermediates shown represents an ensemble of similar structures, and thus a protein follows a general rather than a precise pathway in its transition from the unfolded to the native state. The energy surface for the overall process of protein folding can be visualized as a funnel (Figure 2.60). The wide rim of the funnel represents the wide range of structures accessible to the ensemble of denatured protein molecules. As the free energy of the population of protein molecules decreases, the proteins move down into narrower parts of the funnel and fewer conformations are accessible. At the bottom of the funnel is the folded state with its well-defined conformation. Many paths can lead to this same energy minimum.

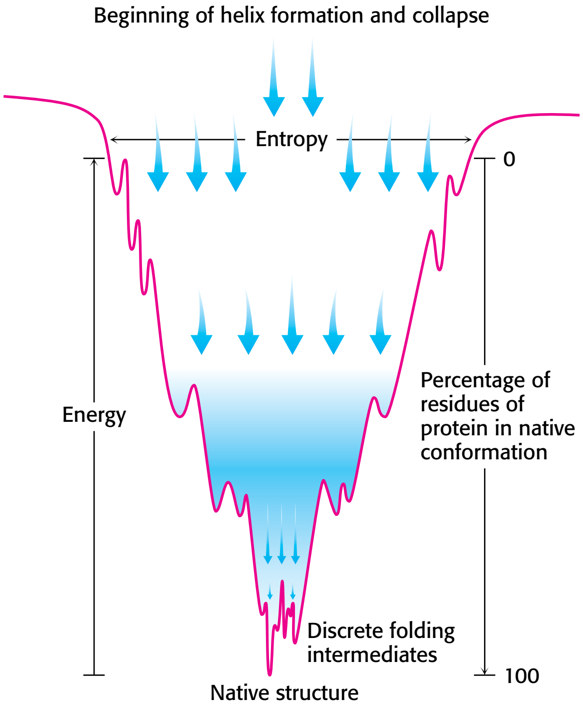

Beginning of helix formation and collapse
Entropy
0
Energy
Percentage of residues of protein in native conformation
Discrete folding intermediates
100
Native structure

FIGURE 2.60 Folding funnel. The folding funnel depicts the thermodynamics of protein folding. The top of the funnel represents all possible denatured conformations—that is, maximal conformational entropy. Depressions on the sides of the funnel represent semistable intermediates that can facilitate or hinder the formation of the native structure, depending on their depth. Secondary structures, such as helices, form and collapse onto one another to initiate folding. [After D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 5th ed. (W. H. Freeman and Company, 2008), p. 143.]

Prediction of three-dimensional structure from sequence remains a great challenge

The prediction of three-dimensional structure from sequence has proved to be extremely difficult. The local sequence appears to determine only between 60 and 70% of the secondary structure; long-range interactions are required to stabilize the full secondary structure and the tertiary structure.

Investigators are exploring two fundamentally different approaches to predicting three-dimensional structure from amino acid sequence. The first is *ab initio* (Latin, "from the beginning") *prediction,* which attempts to predict the folding of an amino acid sequence without prior knowledge about similar sequences in known protein structures. Computer-based calculations are employed that attempt to minimize the free energy of a structure with a given amino acid sequence or to simulate the folding process. The utility of these methods is limited by the vast number of possible conformations, the marginal stability of proteins, and the subtle energetics of weak interactions in aqueous solution. The second approach takes advantage of our growing knowledge of the three-dimensional structures of many proteins. In these *knowledge-based methods,* an amino acid sequence of unknown structure is examined for compatibility with known protein structures or fragments therefrom. If a significant match is detected, the known structure can be used as an initial model. Knowledge-based methods have been a source of many insights into the three-dimensional conformation of proteins of known sequence but unknown structure.

Some proteins are inherently unstructured and can exist in multiple conformations

The discussion of protein folding thus far is based on the paradigm that a given protein amino acid sequence will fold into a particular three-dimensional structure. This paradigm holds well for many proteins. However, it has been known for some time that some proteins can adopt two different structures, only one of which results in protein aggregation and pathological conditions (p. 56). Such alternate structures originating from a unique amino acid sequence were thought to be rare, the exception to the paradigm. Recent work has called into question the universality of the idea that each amino acid sequence gives rise to one structure for certain proteins, even under normal cellular conditions.

Our first example is a class of proteins referred to as *intrinsically unstructured proteins* (IUPs). As the name suggests, these proteins, completely or in part, do not have a discrete three-dimensional structure under physiological conditions. Indeed, an estimated 50% of eukaryotic proteins have at least one unstructured region greater than 30 amino acids in length. Unstructured regions are rich in charged and polar amino acids with few hydrophobic residues. These proteins assume a defined structure on interaction with other proteins. This molecular versatility means that one protein can assume different structures and interact with the different partners, yielding different biochemical functions. IUPs appear to be especially important in signaling and regulatory pathways.

Another class of proteins that do not adhere to the paradigm is *metamorphic proteins*. These proteins appear to exist in an ensemble of structures of approximately equal energy that are in equilibrium. Small molecules or other proteins may bind to different members of the ensemble, resulting in various complexes, each having a different biochemical function. An especially clear example of a metamorphic protein is the chemokine lymphotactin. Chemokines are small signaling proteins in the immune system that bind to receptor proteins on the surface of immune-system cells, instigating an immunological response. Lymphotactin exists in two very different structures that are in equilibrium (Figure 2.61). One structure is a characteristic of chemokines, consisting of a three-stranded β sheet and a carboxyl-terminal helix. This structure binds to its receptor and activates it. The alternative structure is an identical dimer of all β sheets. When in this structure, lymphotactin binds to glycosaminoglycan, a complex carbohydrate (Chapter 11). The biochemical activities of each structure are mutually exclusive: the chemokine structure cannot bind the glycosaminoglycan, and the β-sheet structure cannot activate the receptor. Yet, remarkably, both activities are required for full biological activity of the chemokine.

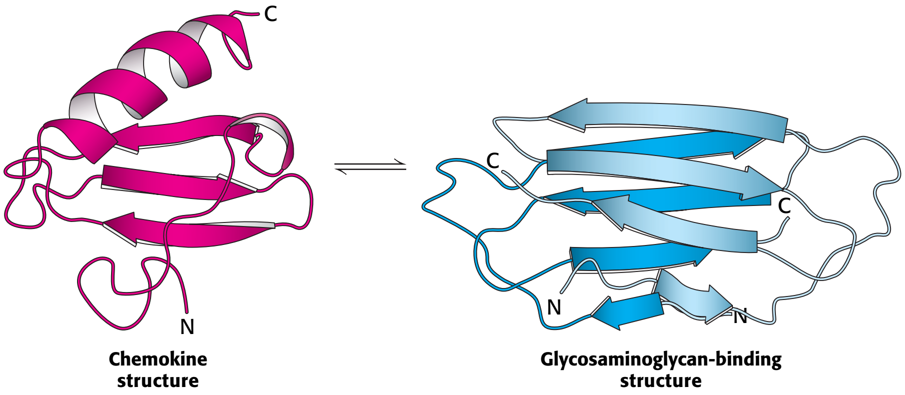


FIGURE 2.61 Lymphotactin exists in two conformations, which are in equilibrium. [R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, and B. F. Volkman. *Proc. Natl. Sci. U.S.A.* 105:5057–5062, 2008, Fig. 2A.]

Note that IUPs and metamorphic proteins effectively expand the protein-encoding capacity of the genome. In some cases, a gene can encode a single protein that has more than one structure and function. These examples also illustrate the dynamic nature of the study of biochemistry and its inherent excitement: even well-established ideas are often subject to modifications.

Protein misfolding and aggregation are associated with some neurological diseases

Understanding protein folding and misfolding is of more than academic interest. A host of diseases, including Alzheimer disease, Parkinson disease, Huntington disease, and transmissible spongiform encephalopathies (prion disease), are associated with improperly folded proteins. All of these diseases result in the deposition of protein aggregates, called *amyloid fibrils* or *plaques*. These diseases are consequently referred to as *amyloidoses*. A common feature of amyloidoses is that normally soluble proteins are converted into insoluble fibrils rich in β sheets. The correctly folded protein is only marginally more stable than the incorrect form. But the incorrect form aggregates, pulling more correct forms into the incorrect form. We will focus on the transmissible spongiform encephalopathies.

One of the great surprises in modern medicine was that certain infectious neurological diseases were found to be transmitted by agents that were similar in size to viruses but consisted only of protein. These diseases include *bovine spongiform encephalopathy* (commonly referred to as *mad cow disease*) and the analogous diseases in other organisms, including *Creutzfeld–Jakob disease* (CJD) in human beings, *scrapie* in sheep, and chronic wasting disease in deer and elk. The agents causing these diseases are termed *prions.* Prions are composed largely, if not exclusively, of a cellular protein called PrP, which is normally present in the brain; its function is still the focus of active research. The infectious prions are aggregated forms of the PrP protein termed PrP^{SC}.

How does the structure of the protein in the aggregated form differ from that of the protein in its normal state in the brain? The normal cellular protein PrP contains extensive regions of α helix and relatively little β strand. The structure of the form of PrP present in infected brains, termed PrP^{SC}, has not yet been determined because of challenges posed by its insoluble and heterogeneous nature. However, a variety of evidence indicates that some parts of the protein that had been in α-helical or turn conformations have been converted into β-strand conformations (Figure 2.62). The β strands of largely planar monomers stack on one another with their side chains tightly interwoven. A side view shows the extensive network of hydrogen bonds between the monomers. These fibrous protein aggregates are often referred to as *amyloid* forms.

With the realization that the infectious agent in prion diseases is an aggregated form of a protein that is already present in the brain, a model for disease transmission emerges (Figure 2.63). Protein aggregates built of abnormal forms of PrP^{SC} act as sites of nucleation to which other PrP molecules attach. Prion diseases can thus be transferred from one individual organism to another through the transfer of an aggregated nucleus, as likely happened in the mad cow disease outbreak in the United Kingdom that emerged in the late 1980s. Cattle fed on animal feed containing material from diseased cows developed the disease in turn.

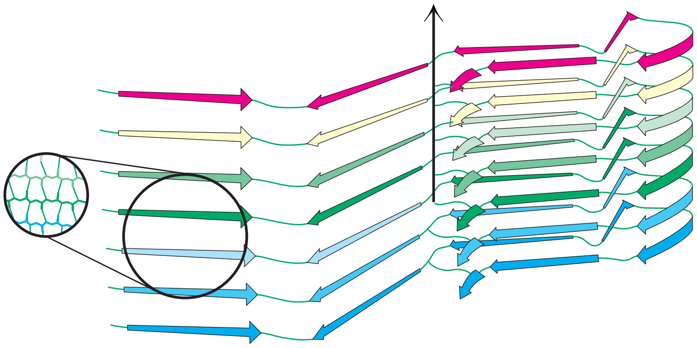


FIGURE 2.62 A model of the human prion protein amyloid. A detailed model of a human prion amyloid fibril deduced from spin labeling and electron paramagnetic resonance (EPR) spectroscopy studies shows that protein aggregation is due to the formation of large parallel β sheets. The black arrow indicates the long axis of the fibril. [N. J. Cobb, F. D. Sönnichsen, H. Mchaourab, and W. K. Surewicz. *Proc. Natl. Acad. Sci. U.S.A.* 104: 18946–18951, 2007, Fig. 4E.]

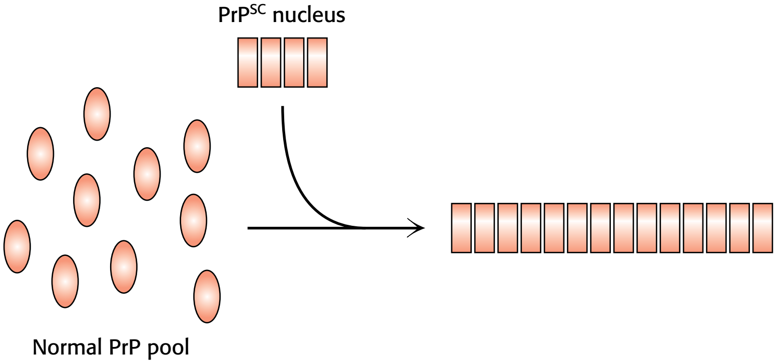


FIGURE 2.63 The protein-only model for prion-disease transmission. A nucleus consisting of proteins in an abnormal conformation grows by the addition of proteins from the normal pool.

Amyloid fibers are also seen in the brains of patients with certain noninfectious neurodegenerative disorders such as Alzheimer and Parkinson diseases. For example, the brains of patients with Alzheimer disease contain protein aggregates called *amyloid plaques* that consist primarily of a single polypeptide termed Aβ. This polypeptide is derived from a cellular protein called *amyloid precursor protein* (APP) through the action of specific proteases. Polypeptide Aβ is prone to form insoluble aggregates. Despite the difficulties posed by the protein's insolubility, a detailed structural model for Aβ has been derived through the use of NMR techniques that can be applied to solids rather than to materials in solution. As expected, the structure is rich in β strands, which come together to form extended parallel β-sheet structures (Figure 2.62).

How do such aggregates lead to the death of the cells that harbor them? The answer is still controversial. One hypothesis is that the large aggregates themselves are not toxic but, instead, smaller aggregates of the same proteins may be the culprits, perhaps damaging cell membranes.

Protein modification and cleavage confer new capabilities

Proteins are able to perform numerous functions that rely solely on the versatility of their 20 amino acids. In addition, many proteins are covalently modified, through the attachment of groups other than amino acids, to augment their functions (Figure 2.64). For example, *acetyl groups* are attached to the amino termini of many proteins, a modification that makes these proteins more resistant to degradation. As discussed earlier (p. 45), the addition of *hydroxyl groups* to many proline residues stabilizes fibers of newly synthesized collagen. The biological significance of this modification is evident in the disease scurvy: a deficiency of vitamin C results in insufficient hydroxylation of collagen, and the abnormal collagen fibers that result are unable to maintain normal tissue strength (Section 27.6). Another specialized amino acid is *γ-carboxyglutamate.* In vitamin K deficiency, insufficient

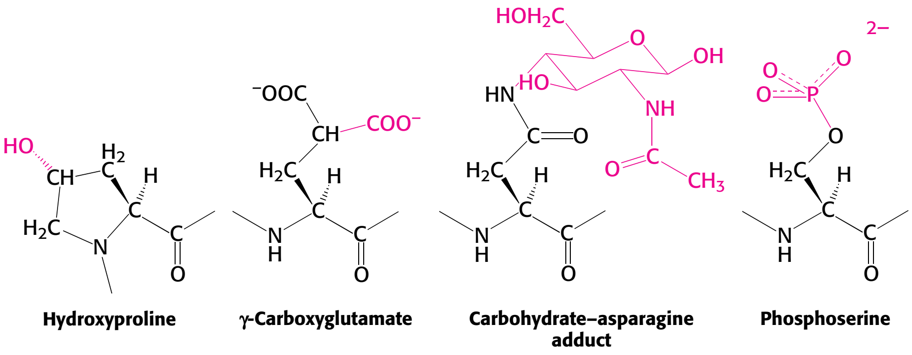


FIGURE 2.64 Finishing touches. Some common and important covalent modifications of amino acid side chains are shown.

carboxylation of glutamate in prothrombin, a clotting protein, can lead to hemorrhage (Section 10.4). Many proteins, especially those that are present on the surfaces of cells or are secreted, acquire *carbohydrate units* on specific asparagine, serine, or threonine residues (Chapter 11). The addition of sugars makes the proteins more hydrophilic and able to participate in interactions with other proteins. Conversely, the addition of a *fatty acid* to an α-amino group or a cysteine sulfhydryl group produces a more hydrophobic protein.

Many hormones, such as epinephrine (adrenaline), alter the activities of enzymes by stimulating the phosphorylation of the hydroxyl amino acids serine and threonine; *phosphoserine* and *phosphothreonine* are the most ubiquitous modified amino acids in proteins. Growth factors such as insulin act by triggering the phosphorylation of the hydroxyl group of tyrosine residues to form *phosphotyrosine*. The phosphoryl groups on these three modified amino acids are readily removed; thus the modified amino acids are able to act as reversible switches in regulating cellular processes. The roles of phosphorylation in signal transduction will be discussed extensively in Chapter 14.

The preceding modifications consist of the addition of special groups to amino acids. Other special groups are generated by chemical rearrangements of side chains and, sometimes, the peptide backbone. For example, the jellyfish *Aequorea victoria* produces green fluorescent protein (GFP), which emits green light when stimulated with blue light. The source of the fluorescence is a group formed by the spontaneous rearrangement and oxidation of the sequence Ser-Tyr-Gly within the center of the protein (Figure 2.65A). Since the discovery

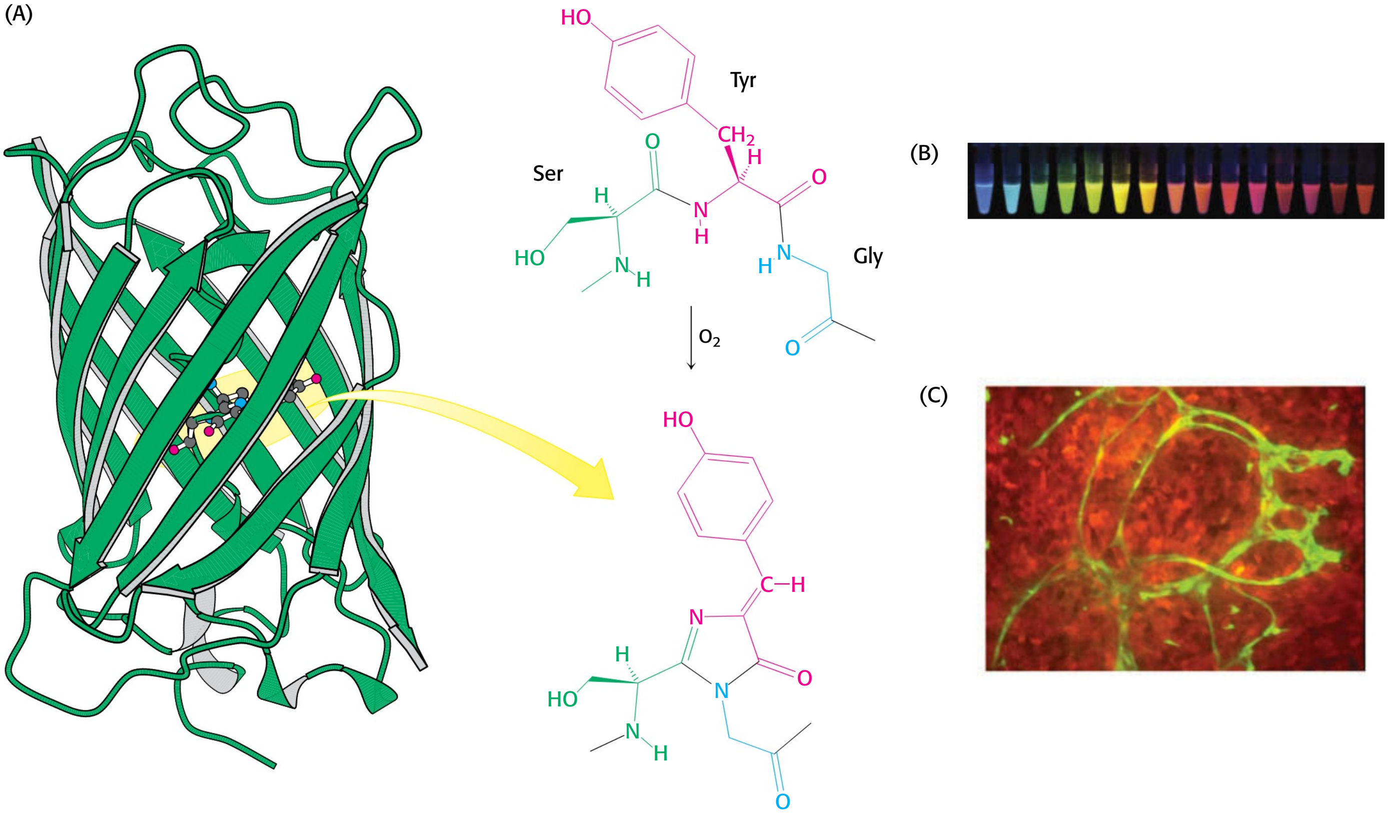


FIGURE 2.65 Chemical rearrangement in GFP. (A) The structure of green fluorescent protein (GFP). The rearrangement and oxidation of the sequence Ser-Tyr-Gly is the source of fluorescence. (B) Mutants of GFP emit light across the visible spectrum. (C) A melanoma cell line engineered to express one of these GFP mutants, red fluorescent protein (RFP), was then injected into a mouse whose blood vessels express GFP. In this fluorescence micrograph, the formation of new blood vessels (green) in the tumor (red) is readily apparent. [(A) Drawn from 1GFL.pdb; (B) R.Y. Tsien. *Integr. Biol.* 2:77–93, 2010, Fig. 12; (C) M. Yang, et al. *Proc. Natl. Acad. Sci. U.S.A.* 100:14259–14262, 2003, Fig. 2B]

of GFP, a number of mutants have been engineered which absorb and emit light across the entire visible spectrum (Figure 2.65B). These proteins are of great utility to researchers as markers within cells (Figure 2.65C).

Finally, many proteins are cleaved and trimmed after synthesis. For example, digestive enzymes are synthesized as inactive precursors that can be stored safely in the pancreas. After release into the intestine, these precursors become activated by peptide-bond cleavage (Section 10.4). In blood clotting, peptide-bond cleavage converts soluble fibrinogen into insoluble fibrin. A number of polypeptide hormones, such as adrenocorticotropic hormone, arise from the splitting of a single large precursor protein. Likewise, many viral proteins are produced by the cleavage of large polyprotein precursors. We shall encounter many more examples of modification and cleavage as essential features of protein formation and function. Indeed, these finishing touches account for much of the versatility, precision, and elegance of protein action and regulation.

SUMMARY

Protein structure can be described at four levels. The primary structure refers to the amino acid sequence. The secondary structure refers to the conformation adopted by local regions of the polypeptide chain. Tertiary structure describes the overall folding of the polypeptide chain. Finally, quaternary structure refers to the specific association of multiple polypeptide chains to form multisubunit complexes.

2.1 Proteins Are Built from a Repertoire of 20 Amino Acids

Proteins are linear polymers of amino acids. Each amino acid consists of a central tetrahedral carbon atom linked to an amino group, a carboxylic acid group, a distinctive side chain, and a hydrogen atom. These tetrahedral centers, with the exception of that of glycine, are chiral; only the L isomer exists in natural proteins. Nearly all natural proteins are constructed from the same set of 20 amino acids. The side chains of these 20 building blocks vary tremendously in size, shape, and the presence of functional groups. They can be grouped as follows: (1) hydrophobic side chains, including the aliphatic amino acids—glycine, alanine, valine, leucine, isoleucine, methionine, and proline—and aromatic side chains—phenylalanine, and tryptophan; (2) polar side chains, including hydroxyl-containing side chains—serine, threonine and tyrosine; the sulfhydryl-containing cysteine; and carboxamide-containing side chains—asparagine and glutamine; (3) basic side chains—lysine, arginine, and histidine; and (4) acidic side chains—aspartic acid and glutamic acid. These groupings are somewhat arbitrary and many other sensible groupings are possible.

2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

The amino acids in a polypeptide are linked by amide bonds formed between the carboxyl group of one amino acid and the amino group of the next. This linkage, called a peptide bond, has several important properties. First, it is resistant to hydrolysis, and so proteins are remarkably stable kinetically. Second, the peptide group is planar because the C—N bond has considerable double-bond character. Third, each peptide bond has both a hydrogen-bond donor (the NH group) and a hydrogen-bond acceptor (the CO group). Hydrogen bonding between these backbone groups is a distinctive feature of protein structure. Finally, the peptide bond is uncharged, which allows proteins to form tightly packed globular structures having significant amounts of the backbone buried within the protein interior. Because they are linear polymers, proteins can be described as sequences of amino acids. Such sequences are written from the amino to the carboxyl terminus.

2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops

Two major elements of secondary structure are the α helix and the β strand. In the α helix, the polypeptide chain twists into a tightly packed rod. Within the helix, the CO group of each amino acid is hydrogen bonded to the NH group of the amino acid four residues farther along the polypeptide chain. In the β strand, the polypeptide chain is nearly fully extended. Two or more β strands connected by NH-to-CO hydrogen bonds come together to form β sheets. The strands in β sheets can be antiparallel, parallel, or mixed.

2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores

The compact, asymmetric structure that individual polypeptides attain is called tertiary structure. The tertiary structures of water-soluble proteins have features in common: (1) an interior formed of amino acids with hydrophobic side chains and (2) a surface formed largely of hydrophilic amino acids that interact with the aqueous environment. The hydrophobic interactions between the interior residues are the driving force for the formation of the tertiary structure of water-soluble proteins. Some proteins that exist in a hydrophobic environment, such as in membranes, display the inverse distribution of hydrophobic and hydrophilic amino acids. In these proteins, the hydrophobic amino acids are on the surface to interact with the environment, whereas the hydrophilic groups are shielded from the environment in the interior of the protein.

2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures

Proteins consisting of more than one polypeptide chain display quaternary structure; each individual polypeptide chain is called a subunit. Quaternary structure can be as simple as two identical subunits or as complex as dozens of different subunits. In most cases, the subunits are held together by noncovalent bonds.

2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

The amino acid sequence determines the three-dimensional structure and, hence, all other properties of a protein. Some proteins can be unfolded completely yet refold efficiently when placed under conditions in which the folded form of the protein is stable. The amino acid sequence of a protein is determined by the sequences of bases in a DNA molecule. This one-dimensional sequence information is extended into the three-dimensional world by the ability of proteins to fold spontaneously. Protein folding is a highly cooperative process; structural intermediates between the unfolded and folded forms do not accumulate.

Some proteins, such as intrinsically unstructured proteins and metamorphic proteins, do not strictly adhere to the one-sequence–one-structure paradigm. Because of this versatility, these proteins expand the protein encoding capacity of the genome.

The versatility of proteins is further enhanced by covalent modifications. Such modifications can incorporate functional groups not present in the 20 amino acids. Other modifications are important to the regulation of protein activity. Through their structural stability, diversity, and chemical reactivity, proteins make possible most of the key processes associated with life.

APPENDIX: Visualizing Molecular Structures II: Proteins

Scientists have developed powerful techniques for the determination of protein structures, as will be considered in Chapter 3. In most cases, these techniques allow the positions of the thousands of atoms within a protein structure to be determined. The final results from such an experiment include the *x*, *y*, and *z* coordinates for each atom in the structure. These coordinate files are compiled in the Protein Data Bank (http://www.pdb.org) from which they can be readily downloaded. These structures comprise thousands or even tens of thousands of atoms. The complexity of proteins with thousands of atoms presents a challenge for the depiction of their structure. Several different types of representations are used to portray proteins, each with its own strengths and weaknesses. The types that you will see most often in this book are space-filling models, ball-and-stick models, backbone models, and ribbon diagrams. Where appropriate, structural features of particular importance or relevance are noted in an illustration's legend.

Space-Filling Models

Space-filling models are the most realistic type of representation. Each atom is shown as a sphere with a size corresponding to the van der Waals radius of the atom (Section 1.3). Bonds are not shown explicitly but are represented by the intersection of the spheres shown when atoms are closer together than the sum of their van der Waals radii. All atoms are shown, including those that make up the backbone and those in the side chains. A space-filling model of lysozyme is depicted in Figure 2.66.

Space-filling models convey a sense of how little open space there is in a protein's structure, which always has many atoms in van der Waals contact with one another. These models are particularly useful in showing conformational changes in a protein from one set of circumstances to another. A disadvantage of space-filling models is that the secondary and tertiary structures of the protein are difficult to see. Thus, these models are not very effective in distinguishing one protein from another—many space-filling models of proteins look very much alike.

Ball-and-Stick Models

Ball-and-stick models are not as realistic as space-filling models. Realistically portrayed atoms occupy more space, determined by their van der Waals radii, than do the atoms depicted in ball-and-stick models. However, the bonding arrangement is easier to see because the bonds are explicitly represented as sticks (Figure 2.67). A ball-and-stick model reveals a complex structure more clearly than a space-filling model does. However, the depiction is so complicated that structural features such as α helices or potential binding sites are difficult to discern.

Because space-filling and ball-and-stick models depict protein structures at the atomic level, the large number of atoms in a complex structure makes it difficult to distinguish the relevant structural features. Thus, representations that are more schematic—such as backbone models and ribbon diagrams—have been developed for the depiction of macromolecular structures. In these representations, most or all atoms are not shown explicitly.

Backbone Models

Backbone models show only the backbone atoms of a polypeptide chain or even only the α-carbon atom of each

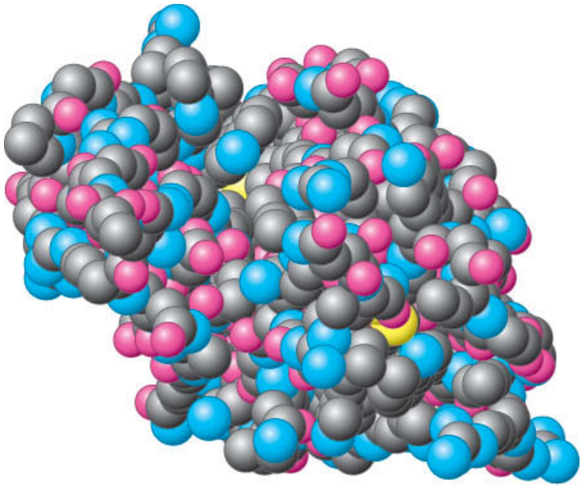


FIGURE 2.66 Space-filling model of lysozyme. *Notice* how tightly packed the atoms are, with little unfilled space. All atoms are shown with the exception of hydrogen atoms. Hydrogen atoms are often omitted because their positions are not readily determined by x-ray crystallographic methods and because their omission somewhat improves the clarity of the structure's depiction.

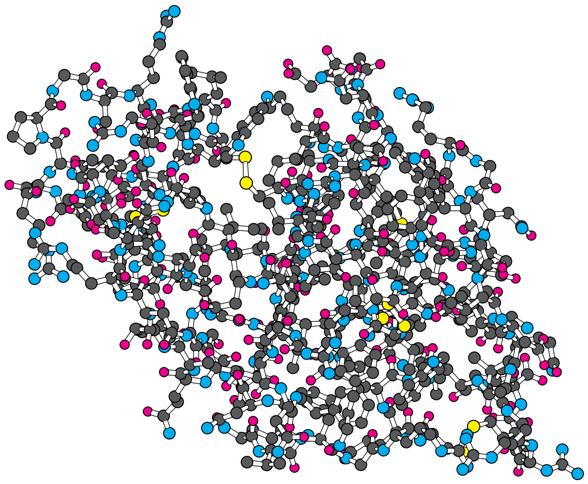


FIGURE 2.67 Ball-and-stick model of lysozyme. Again, hydrogen atoms are omitted.

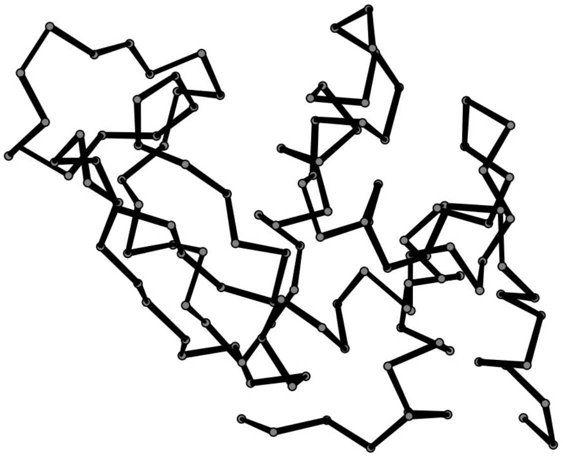


FIGURE 2.68 Backbone model of lysozyme.

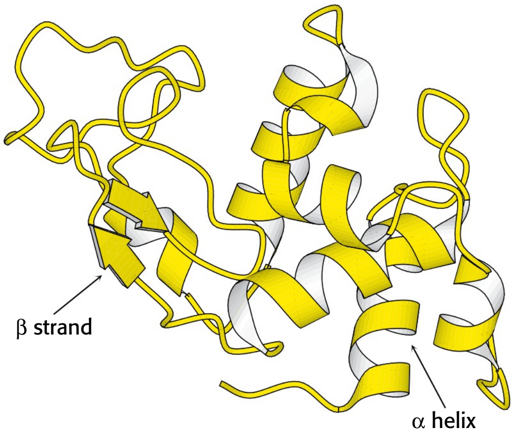


FIGURE 2.69 Ribbon diagram of lysozyme. The α helices are shown as coiled ribbons; β strands are depicted as arrows. More irregular structures are shown as thin tubes.

amino acid. Atoms are linked by lines representing bonds; if only α-carbon atoms are depicted, lines connect α-carbon atoms of amino acids that are adjacent in the amino acid sequence (Figure 2.68). In this book, backbone models show only the lines connecting the α-carbon atoms; other carbon atoms are not depicted.

A backbone model shows the overall course of the polypeptide chain much better than a space-filling or ball-and-stick model does. However, secondary structural elements are still difficult to see.

Ribbon Diagrams

Ribbon diagrams are highly schematic and most commonly used to accent a few dramatic aspects of protein structure, such as the α helix (depicted as a coiled ribbon or a cylinder), the β strand (a broad arrow), and loops (thin tubes), to provide clear views of the folding patterns of proteins (Figure 2.69). The ribbon diagram allows the course of a polypeptide chain to be traced and readily shows the secondary structural elements. Thus, ribbon diagrams of proteins that are related to one another by evolutionary divergence appear similar (Figure 6.15), whereas unrelated proteins are clearly distinct.

In this book, coiled ribbons will be generally used to depict α helices. However, for membrane proteins, which are often quite complex, cylinders will be used rather than coiled ribbons. This convention will also make membrane proteins with their membrane-spanning α helices easy to recognize (Figure 12.18).

Bear in mind that the open appearance of ribbon diagrams is deceptive. As noted earlier, protein structures are tightly packed and have little open space. The openness of ribbon diagrams makes them particularly useful as frameworks in which to highlight additional aspects of protein structure. Active sites, substrates, bonds, and other structural fragments can be included in ball-and-stick or space-filling form within a ribbon diagram (Figure 2.70).

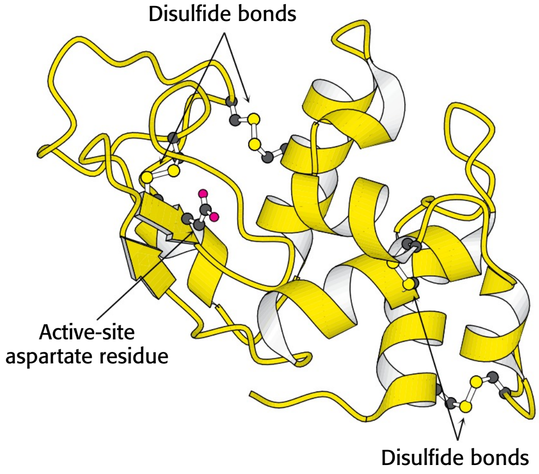


FIGURE 2.70 Ribbon diagram of lysozyme with highlights. Four disulfide bonds and a functionally important aspartate residue are shown in ball-and-stick form.

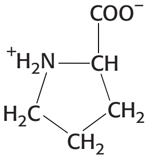
KEY TERMS

side chain (R group) (p. 29)
L amino acid (p. 29)
dipolar ion (zwitterion) (p. 29)
peptide bond (amide bond) (p. 35)
disulfide bond (p. 36)
primary structure (p. 37)
torsion angle (p. 39)
phi (ϕ) angle (p. 39)
psi (ψ) angle (p. 39)
Ramachandran plot (p. 39)
secondary structure (p. 40)
α helix (p. 40)
rise (translation) (p. 40)
β pleated sheet (p. 42)
β strand (p. 42)

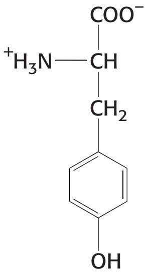
reverse turn (β turn; hairpin turn) (p. 44)
coiled coil (p. 44)
heptad repeat (p. 44)
tertiary structure (p. 46)
motif (supersecondary structure) (p. 48)
domain (p. 48)
subunit (p. 48)
quaternary structure (p. 48)
cooperative transition (p. 52)
intrinsically unstructured protein (IUP) (p. 55)
metamorphic protein (p. 55)
prion (p. 56)

PROBLEMS

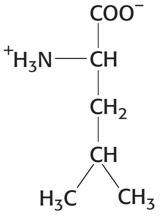
1. *Identify.* Examine the following four amino acids (A–D):



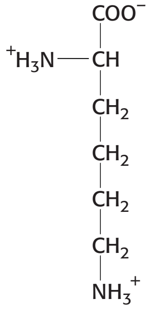
A



B



C



D

What are their names, three-letter abbreviations, and one-letter symbols?

2. *Properties.* In reference to the amino acids shown in Problem 1, which are associated with the following characteristics?

(a) Hydrophobic side chain ______________

(b) Basic side chain ______________

(c) Three ionizable groups ______________

(d) pK_a of approximately 10 in proteins ______________

(e) Modified form of phenylalanine ______________

3. *Match 'em.* Match each amino acid in the left-hand column with the appropriate side-chain type in the right-hand column.

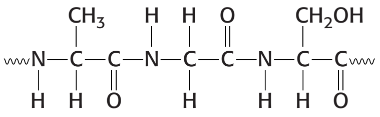
(a) Leu
(b) Glu
(c) Lys
(d) Ser
(e) Cys
(f) Trp

(1) hydroxyl-containing
(2) acidic
(3) basic
(4) sulfur-containing
(5) nonpolar aromatic
(6) nonpolar aliphatic

4. *Solubility.* In each of the following pairs of amino acids, identify which amino acid would be more soluble in water: (a) Ala, Leu; (b) Tyr, Phe; (c) Ser, Ala; (d) Trp, His.

5. *Bonding is good.* Which of the following amino acids have R groups that have hydrogen-bonding potential? Ala, Gly, Ser, Phe, Glu, Tyr, Ile, and Thr.

6. *Name those components.* Examine the segment of a protein shown here.



(a) What three amino acids are present?

(b) Of the three, which is the N-terminal amino acid?

(c) Identify the peptide bonds.

(d) Identify the α-carbon atoms.

7. *Who's charged?* Draw the structure of the dipeptide Gly-His. What is the charge on the peptide at pH 5.5? pH 7.5?

8. *Alphabet soup.* How many different polypeptides of 50 amino acids in length can be made from the 20 common amino acids?

9. *Sweet tooth, but calorie conscious.* Aspartame (NutraSweet), an artificial sweetener, is a dipeptide composed of Asp-Phe in which the carboxyl terminus is modified by the attachment of a methyl group. Draw the structure of Aspartame at pH 7.

10. *Vertebrate proteins?* What is meant by the term *polypeptide backbone?*

11. *Not a sidecar.* Define the term *side chain* in the context of amino acid or protein structure.

12. *One from many.* Differentiate between *amino acid composition* and *amino acid sequence.*

13. *Shape and dimension.* (a) Tropomyosin, a 70-kDa muscle protein, is a two-stranded α-helical coiled coil. Estimate the length of the molecule. (b) Suppose that a 40-residue segment of a protein folds into a two-stranded antiparallel β structure with a 4-residue hairpin turn. What is the longest dimension of this motif?

14. *Contrasting isomers.* Poly-L-leucine in an organic solvent such as dioxane is α helical, whereas poly-L-isoleucine is not. Why do these amino acids with the same number and kinds of atoms have different helix-forming tendencies?

15. *Exceptions to the rule.* Ramachandran plots for two amino acids differ significantly from that shown in Figure 2.23. Which two, and why?

16. *Active again.* A mutation that changes an alanine residue in the interior of a protein to valine is found to lead to a loss of activity. However, activity is regained when a second

mutation at a different position changes an isoleucine residue to glycine. How might this second mutation lead to a restoration of activity?

17. *Exposure issues.* Many of the loops on proteins are composed of hydrophilic amino acids. Why might this be the case?

18. *Shuffle test.* An enzyme that catalyzes disulfide–sulfhydryl exchange reactions, called protein disulfide isomerase (PDI), has been isolated. PDI rapidly converts inactive scrambled ribonuclease into enzymatically active ribonuclease. In contrast, insulin is rapidly inactivated by PDI. What does this important observation imply about the relation between the amino acid sequence of insulin and its three-dimensional structure?

19. *Stretching a target.* A protease is an enzyme that catalyzes the hydrolysis of the peptide bonds of target proteins. How might a protease bind a target protein so that its main chain becomes fully extended in the vicinity of the vulnerable peptide bond?

20. *Often irreplaceable.* Glycine is a highly conserved amino acid residue in the evolution of proteins. Why?

21. *Potential partners.* Identify the groups in a protein that can form hydrogen bonds or electrostatic bonds with an arginine side chain at pH 7.

22. *Permanent waves.* The shape of hair is determined in part by the pattern of disulfide bonds in keratin, its major protein. How can curls be induced?

23. *Location is everything 1.* Most proteins have hydrophilic exteriors and hydrophobic interiors. Would you expect this structure to apply to proteins embedded in the hydrophobic interior of a membrane? Explain.

24. *Location is everything 2.* Proteins that span biological membranes often contain α helices. Given that the insides of membranes are highly hydrophobic (Section 12.2), predict what type of amino acids would be in such an α helix. Why is an α helix particularly suited to existence in the hydrophobic environment of the interior of a membrane?

25. *Neighborhood peer pressure?* Table 2.1 shows the typical pK_a values for ionizable groups in proteins. However, more than 500 pK_a values have been determined for individual groups in folded proteins. Account for this discrepancy.

26. *Greasy patches.* The α and β subunits of hemoglobin bear a remarkable structural similarity to myoglobin. However, in the subunits of hemoglobin, certain residues that are hydrophilic in myoglobin are hydrophobic. Why might this be the case?

27. *Maybe size does matter.* Osteogenesis imperfecta displays a wide range of symptoms, from mild to severe. On the basis of your knowledge of amino acid and collagen structure, propose a biochemical basis for the variety of symptoms.

28. *Issues of stability.* Proteins are quite stable. The lifetime of a peptide bond in aqueous solution is nearly 1000 years. However, the free energy of hydrolysis of proteins is negative and quite large. How can you account for the stability of the peptide bond in light of the fact that hydrolysis releases much energy?

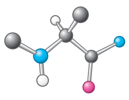
29. *Minor species.* For an amino acid such as alanine, the major species in solution at pH 7 is the zwitterionic form. Assume a pK_a value of 8 for the amino group and a pK_a value of 3 for the carboxylic acid. Estimate the ratio of the concentration of the neutral amino acid species (with the carboxylic acid protonated and the amino group neutral) to that of the zwitterionic species at pH 7 (Section 1.3).

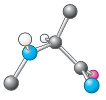
30. *A matter of convention.* All L amino acids have an *S* absolute configuration except L-cysteine, which has the *R* configuration. Explain why L-cysteine is designated as having the *R* absolute configuration.

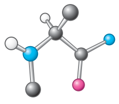
31. *Hidden message.* Translate the following amino acid sequence into one-letter code: Glu-Leu-Val-Ile-Ser-Ile-Ser-Leu-Ile-Val-Ile-Asn-Gly-Ile-Asn-Leu-Ala-Ser-Val-Glu-Gly-Ala-Ser.

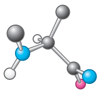
32. *Who goes first?* Would you expect Pro—X peptide bonds to tend to have cis conformations like those of X—Pro bonds? Why or why not?

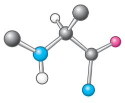
33. *Matching.* For each of the amino acid derivatives shown here (A–E), find the matching set of ϕ and ψ values (a–e).

(A)


(B)


(C)


(D)


(E)


(a) $\phi = 120°$, $\psi = 120°$

(b) $\phi = 180°$, $\psi = 0°$

(c) $\phi = 180°$, $\psi = 180°$

(d) $\phi = 0°$, $\psi = 180°$

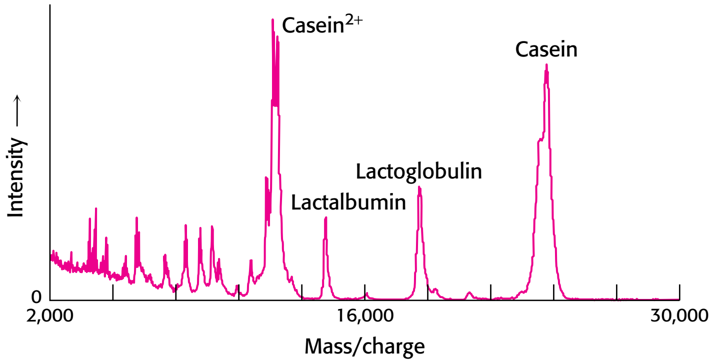
(e) $\phi = -60°$, $\psi = -40°$

34. *Scrambled ribonuclease.* When performing his experiments on protein refolding, Christian Anfinsen obtained a quite different result when reduced ribonuclease was reoxidized while it was still in 8 M urea and the preparation was then dialyzed to remove the urea. Ribonuclease reoxidized in this way had only 1% of the enzymatic activity of the native protein. Why were the outcomes so different when reduced ribonuclease was reoxidized in the presence and absence of urea?

CHAPTER

3

Exploring Proteins and Proteomes



Milk, a source of nourishment for all mammals, is composed, in part, of a variety of proteins. The protein components of milk are revealed by the technique of MALDI–TOF mass spectrometry, which separates molecules on the basis of their mass-to-charge ratio. [(Left) Okea/istockphoto.com. (Right) Courtesy of Dr. Brian Chait.]

Proteins play crucial roles in nearly all biological processes—in catalysis, signal transmission, and structural support. This remarkable range of functions arises from the existence of thousands of proteins, each folded into a distinctive three-dimensional structure that enables it to interact with one or more of a highly diverse array of molecules. A major goal of biochemistry is to determine how amino acid sequences specify the conformations, and hence functions, of proteins. Other goals are to learn how individual proteins bind specific substrates and other molecules, mediate catalysis, and transduce energy and information.

It is often preferable to study a protein of interest after it has been separated from other components within the cell so that the structure and function of this protein can be probed without any confounding effects from contaminants. Hence, the first step in these studies is the purification of the protein of interest. Proteins can be separated from one another on the basis of solubility, size, charge, and binding ability. After a protein has been purified, its amino acid sequence can be determined. Many protein sequences, often deduced from genome sequences, are available in vast sequence databases. If the sequence of a purified protein has been archived in a publicly searchable database, the job of the investigator becomes much easier. The investigator need determine only a small stretch of amino acid sequence of the protein to find its match in the database. Alternatively, such a protein might be identified by matching its mass to those deduced for proteins in the database. Mass spectrometry provides a powerful method for determining the mass and sequence of a protein.

OUTLINE

3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function

3.2 Immunology Provides Important Techniques with Which to Investigate Proteins

3.3 Mass Spectrometry Is a Powerful Technique for the Identification of Peptides and Proteins

3.4 Peptides Can Be Synthesized by Automated Solid-Phase Methods

3.5 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy

After a protein has been purified and its identity confirmed, the challenge remains to determine its function within a physiologically relevant context. Antibodies are choice probes for locating proteins in vivo and measuring their quantities. Monoclonal antibodies, able to recognize specific proteins, can be obtained in large amounts and used to detect and quantify the protein both in isolation and in cells. Peptides and proteins can be chemically synthesized, providing tools for research and, in some cases, highly pure material for use as drugs. Finally, x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the principal techniques for elucidating three-dimensional structure, the key determinant of function.

The exploration of proteins by this array of physical and chemical techniques has greatly enriched our understanding of the molecular basis of life. These techniques make it possible to tackle some of the most challenging questions of biology in molecular terms.

The proteome is the functional representation of the genome

As will be discussed in Chapter 5, the complete DNA base sequences, or *genomes,* of many organisms are now available. For example, the roundworm *Caenorhabditis elegans* has a genome of 97 million bases and about 19,000 protein-encoding genes, whereas that of the fruit fly *Drosophila melanogaster* contains 180 million bases and about 14,000 genes. The completely sequenced human genome contains 3 billion bases and about 23,000 genes. However, these genomes are simply inventories of the genes that *could* be expressed within a cell under specific conditions. Only a subset of the proteins encoded by these genes will actually be present in a given biological context. The *proteome*—derived from *prote*ins expressed by the gen*ome*—of an organism signifies a more complex level of information content, encompassing the types, functions, and interactions of proteins within its biological environment.

The proteome is not a fixed characteristic of the cell. Because it represents the functional expression of information, it varies with cell type, developmental stage, and environmental conditions, such as the presence of hormones. The proteome is much larger than the genome because almost all gene products are proteins that can be chemically modified in a variety of ways. Furthermore, these proteins do not exist in isolation; they often interact with one another to form complexes with specific functional properties. Whereas the genome is "hard wired," the proteome is highly dynamic. An understanding of the proteome is acquired by investigating, characterizing, and cataloging proteins. In some, but not all, cases, this process begins by separating a particular protein from all other biomolecules in the cell.

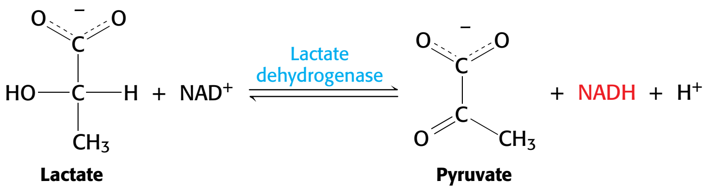
3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function

An adage of biochemistry is "Never waste pure thoughts on an impure protein." Starting from pure proteins, we can determine amino acid sequences and investigate biochemical functions. From the amino acid sequences, we can map evolutionary relationships between proteins in diverse organisms (Chapter 6). By using crystals grown from pure protein, we can obtain x-ray data that will provide us with a picture of the protein's tertiary structure—the shape that determines function.

The assay: How do we recognize the protein that we are looking for?

Purification should yield a sample containing only one type of molecule—the protein in which the biochemist is interested. This protein sample may be only a fraction of 1% of the starting material, whether that starting material consists of one type of cell in culture or a particular organ from a plant or animal. How is the biochemist able to isolate a particular protein from a complex mixture of proteins?

A protein can be purified by subjecting the impure mixture of the starting material to a series of separations based on physical properties such as size and charge. To monitor the success of this purification, the biochemist needs a test, called an *assay,* for some unique identifying property of the protein. A positive result on the assay indicates that the protein is present. Although assay development can be a challenging task, the more specific the assay, the more effective the purification. For enzymes, which are protein catalysts (Chapter 8), the assay usually measures *enzyme activity*—that is, the ability of the enzyme to promote a particular chemical reaction. This activity is often measured indirectly. Consider the enzyme lactate dehydrogenase, which catalyzes the following reaction in the synthesis of glucose:



Reduced nicotinamide adenine dinucleotide (NADH, Figure 15.13) absorbs light at 340 nm, whereas oxidized nicotinamide adenine dinucleotide (NAD^+) does not. Consequently, we can follow the progress of the reaction by examining how much light-absorbing ability is developed by a sample in a given period of time—for instance, within 1 minute after the addition of the enzyme. Our assay for enzyme activity during the purification of lactate dehydrogenase is thus the increase in the absorbance of light at 340 nm observed in 1 minute.

To analyze how our purification scheme is working, we need one additional piece of information—the amount of protein present in the mixture being assayed. There are various rapid and reasonably accurate means of determining protein concentration. With these two experimentally determined numbers—enzyme activity and protein concentration—we then calculate the *specific activity,* the ratio of enzyme activity to the amount of protein in the mixture. Ideally, the specific activity will rise as the purification proceeds and the protein mixture will contain the protein of interest to a greater extent. In essence, the overall goal of the purification is to maximize the specific activity. For a pure enzyme, the specific activity will have a constant value.

Proteins must be released from the cell to be purified

Having found an assay and chosen a source of protein, we now fractionate the cell into components and determine which component is enriched in the protein of interest. In the first step, a *homogenate* is formed by disrupting the cell membrane, and the mixture is fractionated by centrifugation, yielding a dense pellet of heavy material at the bottom of the centrifuge tube and a lighter supernatant above (Figure 3.1). The supernatant is again centrifuged

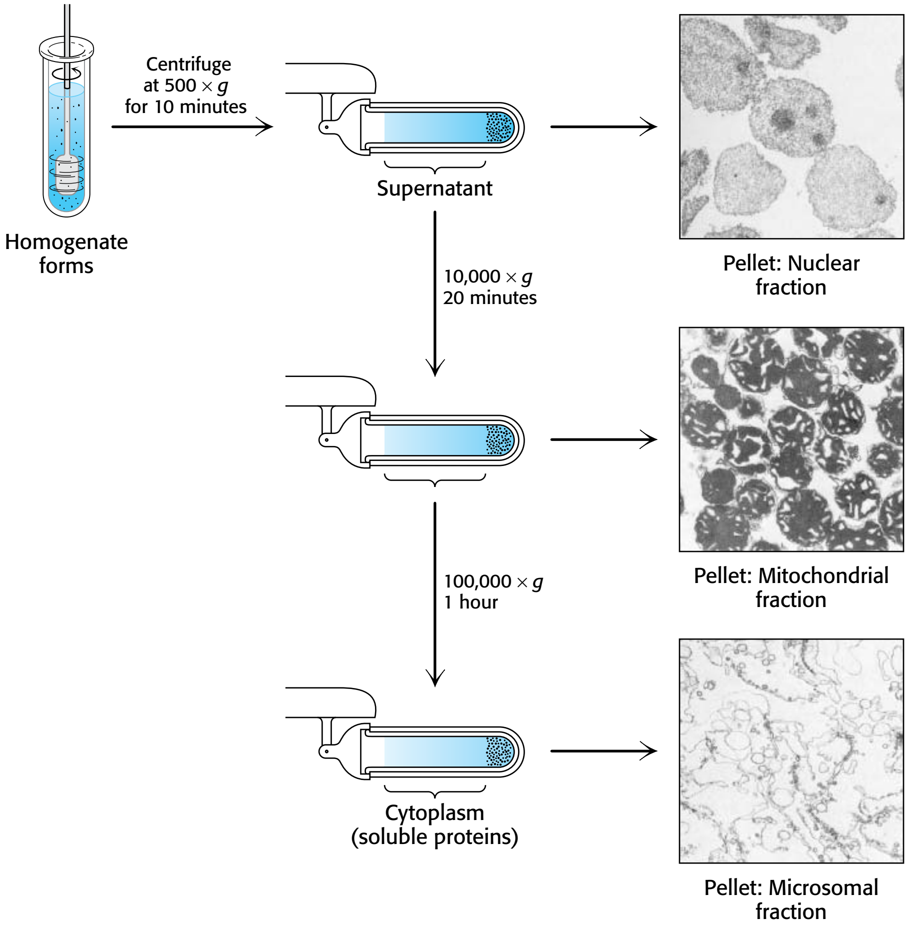

Centrifuge at 500 × g for 10 minutes
Homogenate forms
Supernatant
Pellet: Nuclear fraction
10,000 × g 20 minutes
Pellet: Mitochondrial fraction
100,000 × g 1 hour
Cytoplasm (soluble proteins)
Pellet: Microsomal fraction

FIGURE 3.1 Differential centrifugation. Cells are disrupted in a homogenizer and the resulting mixture, called the homogenate, is centrifuged in a step-by-step fashion of increasing centrifugal force. The denser material will form a pellet at lower centrifugal force than will the less-dense material. The isolated fractions can be used for further purification. [Photographs courtesy of Dr. S. Fleischer and Dr. B. Fleischer.]

at a greater force to yield yet another pellet and supernatant. This procedure, called *differential centrifugation,* yields several fractions of decreasing density, each still containing hundreds of different proteins. The fractions are each separately assayed for the desired activity. Usually, one fraction will be enriched for such activity, and it then serves as the source of material to which more-discriminating purification techniques are applied.

Proteins can be purified according to solubility, size, charge, and binding affinity

Several thousand proteins have been purified in active form on the basis of such characteristics as *solubility, size, charge,* and *specific binding affinity*. Usually, protein mixtures are subjected to a series of separations, each based on a different property. At each step in the purification, the preparation is assayed and its specific activity is determined. A variety of purification techniques are available.

Salting out. Most proteins are less soluble at high salt concentrations, an effect called *salting out*. The salt concentration at which a protein precipitates differs from one protein to another. Hence, salting out can be used to fractionate proteins. For example, 0.8 M ammonium sulfate precipitates fibrinogen, a blood-clotting protein, whereas a concentration of 2.4 M is needed to precipitate serum albumin. Salting out is also useful for concentrating dilute solutions of proteins, including active fractions obtained from other purification steps. Dialysis can be used to remove the salt if necessary.

Dialysis. Proteins can be separated from small molecules such as salt by *dialysis* through a semipermeable membrane, such as a cellulose membrane with pores (Figure 3.2). The protein mixture is placed inside the dialysis bag, which is then submerged in a buffer solution that is devoid of the small molecules to be separated away. Molecules having dimensions significantly greater than the pore diameter are retained inside the dialysis bag. Smaller molecules and ions capable of passing through the pores of the membrane diffuse down their concentration gradients and emerge in the solution outside the bag. This technique is useful for removing a salt or other small molecule from a cell fractionate, but it will not distinguish between proteins effectively.

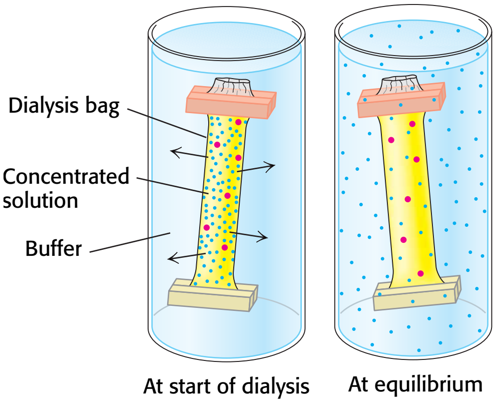
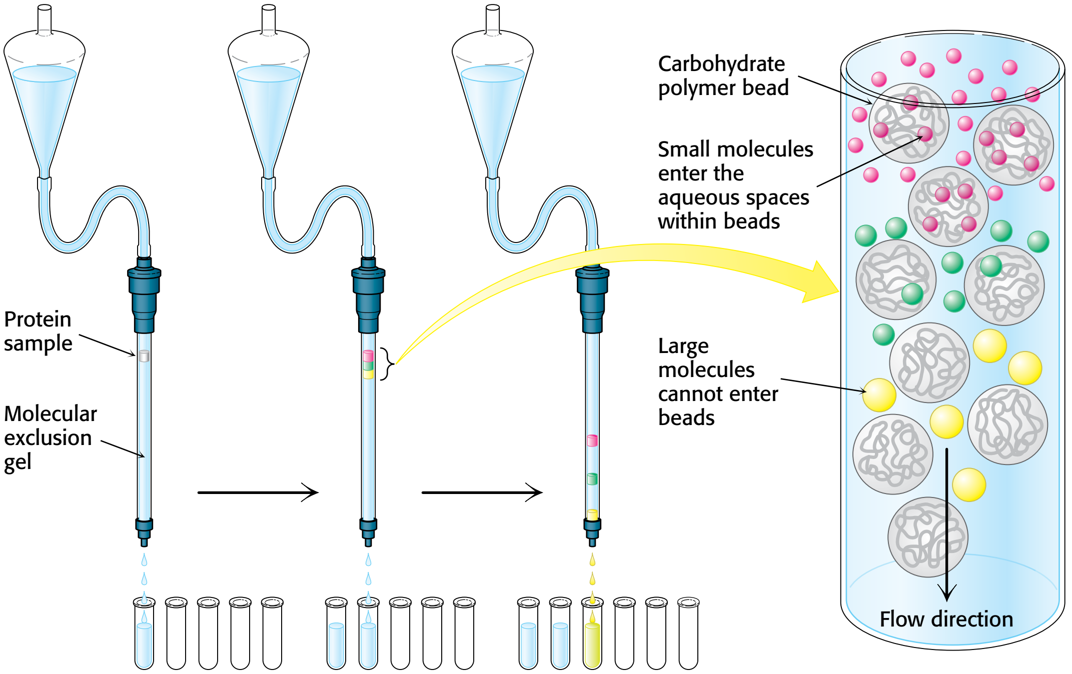


FIGURE 3.2 **Dialysis.** Protein molecules (red) are retained within the dialysis bag, whereas small molecules (blue) diffuse down their concentration gradient into the surrounding medium.

Gel-filtration chromatography. More-discriminating separations on the basis of size can be achieved by the technique of *gel-filtration chromatography,* also known as molecular exclusion chromatography (Figure 3.3). The sample is applied to the top of a column consisting of porous beads made of an insoluble but highly hydrated polymer such as dextran or agarose (which are carbohydrates) or polyacrylamide. Sephadex, Sepharose, and Biogel are commonly used commercial preparations of these beads, which are typically 100 μm (0.1 mm) in diameter. Small molecules can enter these beads, but large ones cannot. The result is that small molecules are distributed in the aqueous solution both inside the beads and between them, whereas large molecules are located only in the solution between the beads. *Large molecules flow more rapidly through this column and emerge first because a smaller volume is accessible to them.* Molecules of medium size occasionally enter the beads and will flow from the column at an intermediate position, while small molecules, which take a longer, tortuous path, will exit last.

Ion-exchange chromatography. To obtain a protein of high purity, one chromatography step is usually not sufficient, because other proteins in the crude mixture will likely co-elute with the desired material. Additional



Protein sample
Molecular exclusion gel
Carbohydrate polymer bead
Small molecules enter the aqueous spaces within beads
Large molecules cannot enter beads
Flow direction

FIGURE 3.3 **Gel-filtration chromatography.** A mixture of proteins in a small volume is applied to a column filled with porous beads. Because large proteins cannot enter the internal volume of the beads, they emerge sooner than do small ones.

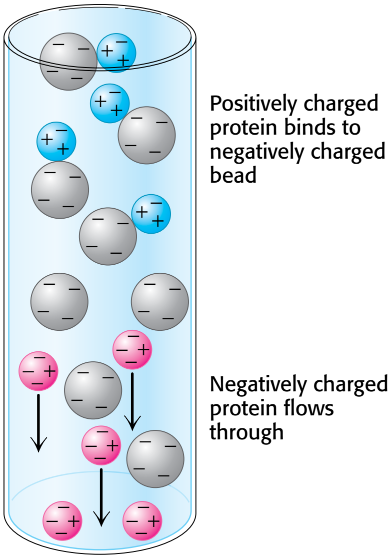
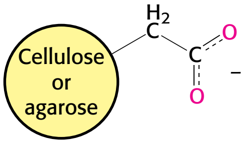
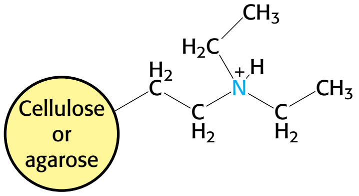

Positively charged protein binds to negatively charged bead
Negatively charged protein flows through

FIGURE 3.4 **Ion-exchange chromatography.** This technique separates proteins mainly according to their net charge.

purity can be achieved by performing sequential separations that are based on distinct molecular properties. For example, in addition to size, proteins can be separated on the basis of their net charge by *ion-exchange chromatography.* If a protein has a net positive charge at pH 7, it will usually bind to a column of beads containing carboxylate groups, whereas a negatively charged protein will not (Figure 3.4). The bound protein can then be eluted (released) by increasing the concentration of sodium chloride or another salt in the eluting buffer; sodium ions compete with positively charged groups on the protein for binding to the column. Proteins that have a low density of net positive charge will tend to emerge first, followed by those having a higher charge density. This procedure is also referred to as *cation exchange* to indicate that positively charged groups will bind to the anionic beads. Positively charged proteins (cationic proteins) can be separated by chromatography on negatively charged carboxymethylcellulose (CM-cellulose) columns. Conversely, negatively charged proteins (anionic proteins) can be separated by *anion exchange* on positively charged diethylaminoethylcellulose (DEAE-cellulose) columns.



Carboxymethyl (CM) group
(ionized form)



Diethylaminoethyl (DEAE) group
(protonated form)

Affinity chromatography. *Affinity chromatography* is another powerful means of purifying proteins that is highly selective for the protein of interest. This technique takes advantage of the high affinity of many proteins for specific chemical groups. For example, the plant protein concanavalin A is a carbohydrate-binding protein, or lectin (Section 11.4), that has affinity for glucose. When a crude extract is passed through a column of beads containing covalently attached glucose residues, concanavalin A binds to the beads, whereas most other proteins do not (Figure 3.5). The bound concanavalin A can then be released from the column by adding a concentrated solution of glucose. The glucose in solution displaces the column-attached glucose residues from binding sites on concanavalin A. Affinity chromatography is a powerful means of isolating transcription factors—proteins that regulate gene expression by binding to specific DNA sequences. A protein mixture is passed through a column containing specific DNA sequences attached to a matrix; proteins with a high affinity for the sequence will bind and be retained. In this instance, the transcription factor is released by washing with a solution containing a high concentration of salt.

In general, affinity chromatography can be effectively used to isolate a protein that recognizes group X by (1) covalently attaching X or a derivative of it to a column; (2) adding a mixture of proteins to this column, which is then washed with buffer to remove unbound proteins; and (3) eluting the desired protein by adding a high concentration of a soluble form of X or altering the conditions to decrease binding affinity. Affinity chromatography is most effective when the interaction of the protein and the molecule that is used as the bait is highly specific.

Affinity chromatography can be used to isolate proteins expressed from cloned genes (Section 5.2). Extra amino acids are encoded in the cloned gene

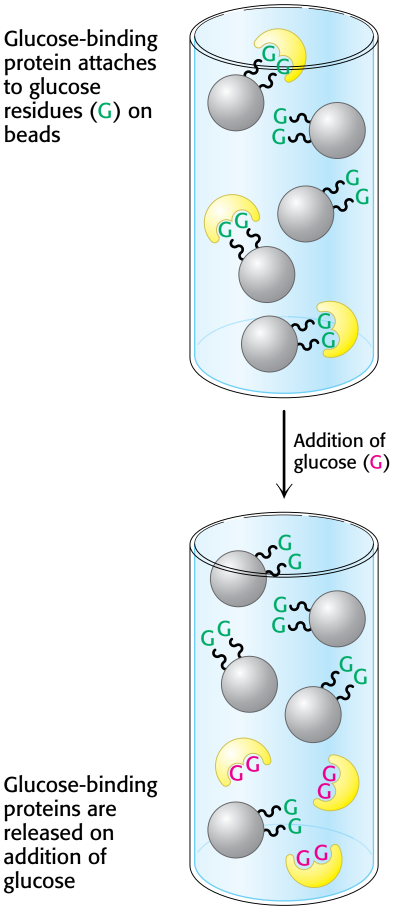

Glucose-binding protein attaches to glucose residues (G) on beads
Addition of glucose (G)
Glucose-binding proteins are released on addition of glucose

FIGURE 3.5 **Affinity chromatography.** Affinity chromatography of concanavalin A (shown in yellow) using a solid support containing covalently attached glucose residues (G).

that, when expressed, serve as an affinity tag that can be readily trapped. For example, repeats of the codon for histidine may be added such that the expressed protein has a string of histidine residues (called a *His tag*) on one end. The tagged proteins are then passed through a column of beads containing covalently attached, immobilized nickel(II) or other metal ions. The His tags bind tightly to the immobilized metal ions, binding the desired protein, while other proteins flow through the column. The protein can then be eluted from the column by the addition of imidazole or some other chemical that binds to the metal ions and displaces the protein.

High-performance liquid chromatography. A technique called *high-performance liquid chromatography* (HPLC) is an enhanced version of the column techniques already discussed. The column materials are much more finely divided and, as a consequence, possess more interaction sites and thus greater resolving power. Because the column is made of finer material, pressure must be applied to the column to obtain adequate flow rates. The net result is both high resolution and rapid separation. In a typical HPLC setup, a detector that monitors the absorbance of the eluate at a particular wavelength is placed immediately after the column. In the sample HPLC elution profile shown in Figure 3.6, proteins are detected by setting the detector to 220 nm (the characteristic absorbance wavelength of the peptide bond). In a short span of 10 minutes, a number of sharp peaks representing individual proteins can be readily identified.

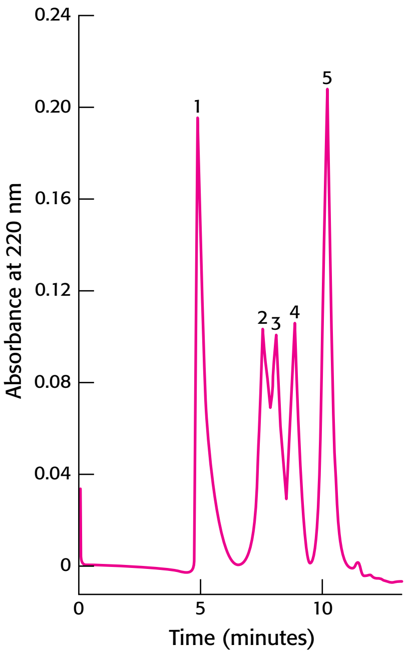


FIGURE 3.6 High-performance liquid chromatography (HPLC). Gel filtration by HPLC clearly defines the individual proteins because of its greater resolving power: (1) thyroglobulin (669 kDa), (2) catalase (232 kDa), (3) bovine serum albumin (67 kDa), (4) ovalbumin (43 kDa), and (5) ribonuclease (13.4 kDa). [Data from K. J. Wilson and T. D. Schlabach. In *Current Protocols in Molecular Biology,* vol. 2, suppl. 41, F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds. (Wiley, 1998), p. 10.14.1.]

Proteins can be separated by gel electrophoresis and displayed

How can we tell that a purification scheme is effective? One way is to ascertain that the specific activity rises with each purification step. Another is to determine that the number of different proteins in each sample declines at each step. The technique of electrophoresis makes the latter method possible.

Gel electrophoresis. A molecule with a net charge will move in an electric field. This phenomenon, termed *electrophoresis,* offers a powerful means of separating proteins and other macromolecules, such as DNA and RNA. The velocity of migration (v) of a protein (or any molecule) in an electric field depends on the electric field strength (E), the net charge on the protein (z), and the frictional coefficient (f).

$$v = Ez/f \tag{1}$$

The electric force Ez driving the charged molecule toward the oppositely charged electrode is opposed by the viscous drag fv arising from friction between the moving molecule and the medium. The frictional coefficient f depends on both the mass and shape of the migrating molecule and the viscosity (η) of the medium. For a sphere of radius r,

$$f = 6\pi\eta r \tag{2}$$

Electrophoretic separations are nearly always carried out in porous gels (or on solid supports such as paper) because the gel serves as a molecular sieve that enhances separation (Figure 3.7). Molecules that are small compared with the pores in the gel readily move through the gel, whereas molecules much larger than the pores are almost immobile. Intermediate-size molecules move through the gel with various degrees of facility. The electric field is applied such that proteins migrate from the negative to the positive electrodes, typically from top to bottom. Electrophoresis is performed in a thin, vertical slab of polyacrylamide gel. Polyacrylamide gels are choice

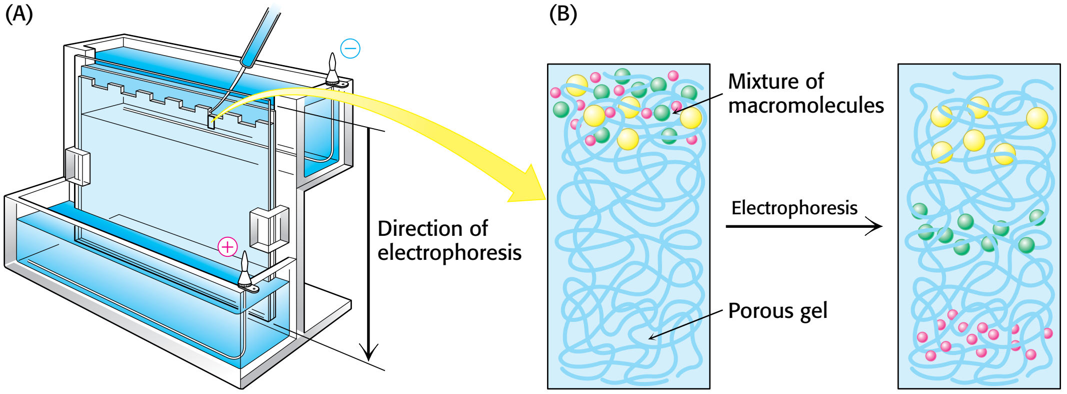

(A)
(B)
Direction of electrophoresis
Mixture of macromolecules
Electrophoresis
Porous gel

FIGURE 3.7 Polyacrylamide gel electrophoresis. (A) Gel-electrophoresis apparatus. Typically, several samples undergo electrophoresis on one flat polyacrylamide gel. A microliter pipette is used to place solutions of proteins in the wells of the slab. A cover is then placed over the gel chamber and voltage is applied. The negatively charged SDS (sodium dodecyl sulfate)–protein complexes migrate in the direction of the anode, at the bottom of the gel. (B) The sieving action of a porous polyacrylamide gel separates proteins according to size, with the smallest moving most rapidly.

supporting media for electrophoresis because they are chemically inert and readily formed by the polymerization of acrylamide with a small amount of the cross-linking agent methylenebisacrylamide to make a three-dimensional mesh (Figure 3.8). Electrophoresis is distinct from gel filtration in that, because of the electric field, all of the molecules, regardless of size, are forced to move through the same matrix.

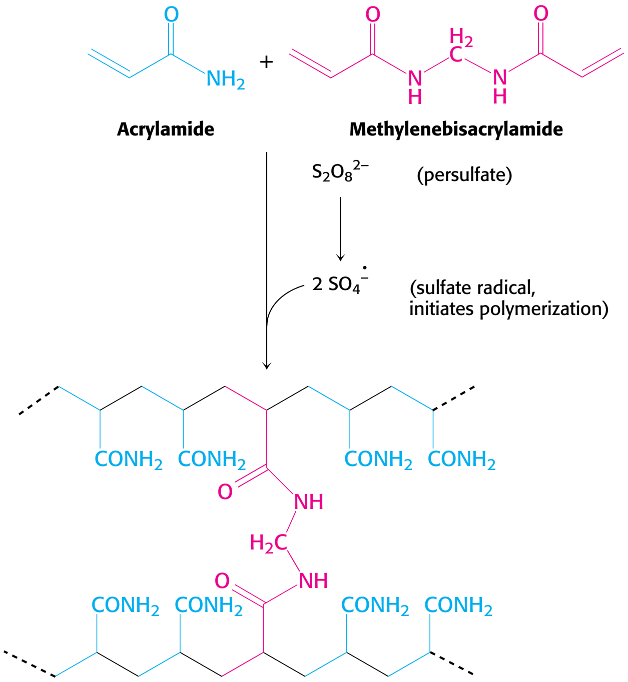
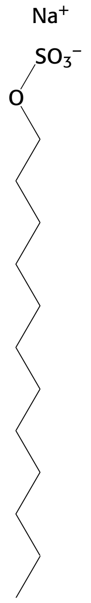

Acrylamide
Methylenebisacrylamide
$S_2O_8^{2-}$ (persulfate)
$2\ SO_4^{\cdot -}$ (sulfate radical, initiates polymerization)
$CONH_2$
NH_2
NH
H_2C
O

FIGURE 3.8 Formation of a polyacrylamide gel. A three-dimensional mesh is formed by copolymerizing activated monomer (blue) and cross-linker (red).

Proteins can be separated largely on the basis of mass by electrophoresis in a polyacrylamide gel under denaturing conditions. The mixture of proteins is first dissolved in a solution of sodium dodecyl sulfate (SDS), an anionic detergent that disrupts nearly all noncovalent interactions in native proteins. β-Mercaptoethanol (2-thioethanol) or dithiothreitol is added to reduce disulfide bonds. Anions of SDS bind to main chains at a ratio of about one SDS anion for every two amino acid residues. The negative charge acquired on binding SDS is usually much greater than the charge on the native protein; the contribution of the protein to the total charge of the



Sodium dodecyl sulfate (SDS)

SDS–protein complex is thus rendered insignificant. As a result, this complex of SDS with a denatured protein has a large net negative charge that is roughly proportional to the mass of the protein. The SDS–protein complexes are then subjected to electrophoresis. When the electrophoresis is complete, the proteins in the gel can be visualized by staining them with silver nitrate or a dye such as Coomassie blue, which reveals a series of bands (Figure 3.9). Radioactive labels, if they have been incorporated into proteins, can be detected by placing a sheet of x-ray film over the gel, a procedure called *autoradiography*.

Small proteins move rapidly through the gel, whereas large proteins stay at the top, near the point of application of the mixture. The mobility of most polypeptide chains under these conditions is linearly proportional to the logarithm of their mass (Figure 3.10). Some carbohydrate-rich proteins and membrane proteins do not obey this empirical relation, however. SDS–polyacrylamide gel electrophoresis (often referred to as SDS-PAGE) is rapid, sensitive, and capable of a high degree of resolution. As little as 0.1 μg (~2 pmol) of a protein gives a distinct band when stained with Coomassie blue, and even less (~0.02 μg) can be detected with a silver stain. Proteins that differ in mass by about 2% (e.g., 50 and 51 kDa, arising from a difference of about 10 amino acids) can usually be distinguished with SDS-PAGE.

We can examine the efficacy of our purification scheme by analyzing a part of each fraction by electrophoresis. The initial fractions will display dozens to hundreds of proteins. As the purification progresses, the number of bands will diminish, and the prominence of one of the bands should increase. This band should correspond to the protein of interest.

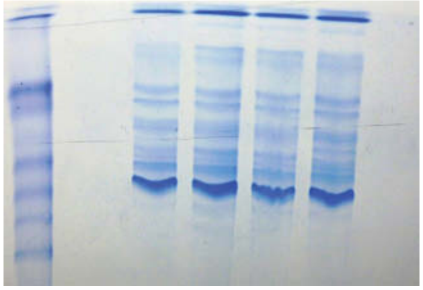


FIGURE 3.9 Staining of proteins after electrophoresis. Mixtures of proteins from cellular extracts subjected to electrophoresis on an SDS–polyacrylamide gel can be visualized by staining with Coomassie blue. The first lane contains a mixture of proteins of known molecular weights, which can be used to estimate the sizes of the bands in the samples. [©Dr. Robert Farrell.]

Isoelectric focusing. Proteins can also be separated electrophoretically on the basis of their relative contents of acidic and basic residues. The *isoelectric point* (pI) of a protein is the pH at which its net charge is zero. At this pH, its electrophoretic mobility is zero because z in equation 1 is equal to zero. For example, the pI of cytochrome *c*, a highly basic electron-transport protein, is 10.6, whereas that of serum albumin, an acidic protein in blood, is 4.8. Suppose that a mixture of proteins undergoes electrophoresis in a pH gradient in a gel in the absence of SDS. Each protein will move until it reaches a position in the gel at which the pH is equal to the pI of the protein. This method of separating proteins according to their isoelectric point is called *isoelectric focusing*. The pH gradient in the gel is formed first by subjecting a mixture of *polyampholytes* (small multi-charged polymers) having many different pI values to electrophoresis. Isoelectric focusing can readily resolve proteins that differ in pI by as little as 0.01, which means that proteins differing by one net charge can be separated (Figure 3.11).

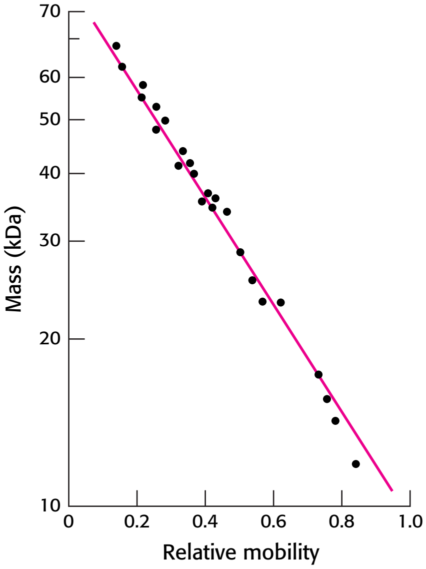


FIGURE 3.10 Electrophoresis can determine mass. The electrophoretic mobility of many proteins in SDS–polyacrylamide gels is inversely proportional to the logarithm of their mass. [Data from K. Weber and M. Osborn, *The Proteins,* vol. 1, 3d ed. (Academic Press, 1975), p. 179.]

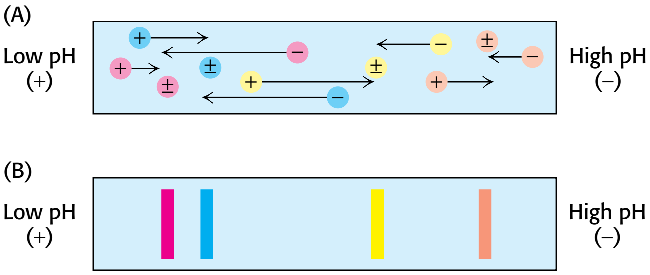


FIGURE 3.11 The principle of isoelectric focusing. A pH gradient is established in a gel before loading the sample. (A) Each protein, represented by the different colored circles, will possess a net positive charge in the regions of the gel where the pH is lower than its respective pI value and a net negative charge where the pH is greater than its pI. When voltage is applied to the gel, each protein will migrate to its pI, the location at which it has no net charge. (B) The proteins form bands that can be excised and used for further experimentation.

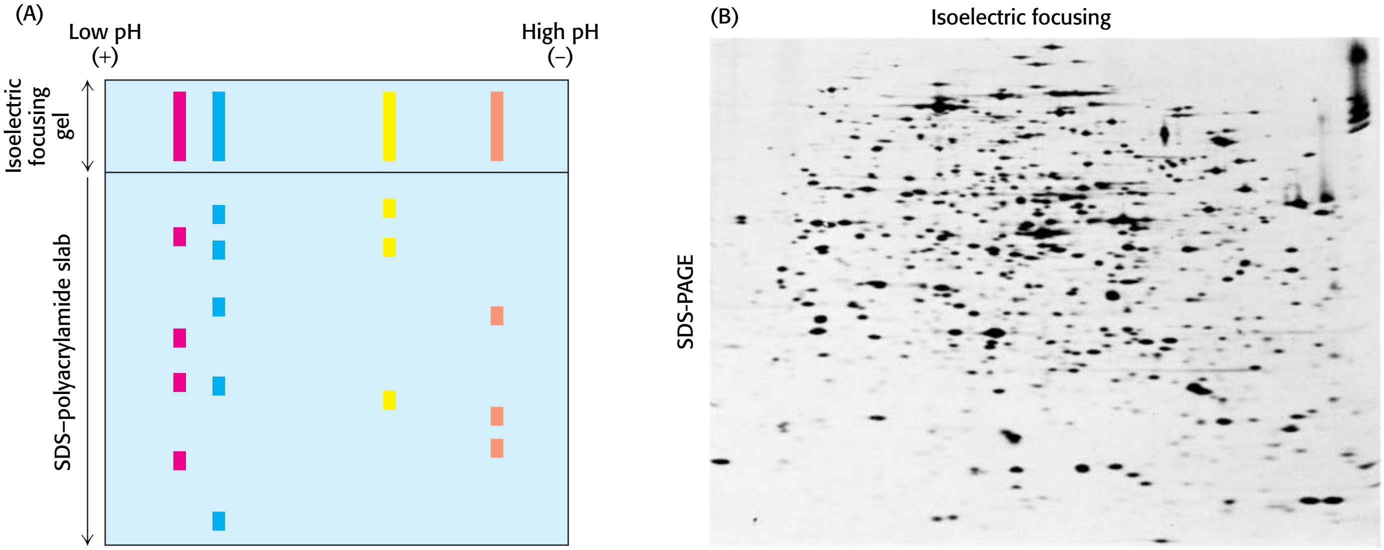


FIGURE 3.12 Two-dimensional gel electrophoresis. (A) A protein sample is initially fractionated in one dimension by isoelectric focusing as described in Figure 3.11. The isoelectric focusing gel is then attached to an SDS–polyacrylamide gel, and electrophoresis is performed in the second dimension, perpendicular to the original separation. Proteins with the same pI are now separated on the basis of mass. (B) Proteins from *E. coli* were separated by two-dimensional gel electrophoresis, resolving more than a thousand different proteins. [(B) Courtesy of Dr. Patrick H. O'Farrell.]

Two-dimensional electrophoresis. Isoelectric focusing can be combined with SDS-PAGE to obtain very high resolution separations. A single sample is first subjected to isoelectric focusing. This single-lane gel is then placed horizontally on top of an SDS–polyacrylamide slab. The proteins are thus spread across the top of the polyacrylamide gel according to how far they migrated during isoelectric focusing. They then undergo electrophoresis again in a perpendicular direction (vertically) to yield a two-dimensional pattern of spots. In such a gel, proteins have been separated in the horizontal direction on the basis of isoelectric point and in the vertical direction on the basis of mass. Remarkably, more than a thousand different proteins in the bacterium *Escherichia coli* can be resolved in a single experiment by two-dimensional electrophoresis (Figure 3.12).

Proteins isolated from cells under different physiological conditions can be subjected to two-dimensional electrophoresis. The intensities of individual spots on the gels can then be compared, which indicates that the concentrations of specific proteins have changed in response to the physiological state (Figure 3.13). How can we discover the identity of a protein that is showing such responses? Although many proteins are displayed on a two-dimensional gel, they are not identified. It is now possible to identify proteins by coupling two-dimensional gel electrophoresis with mass

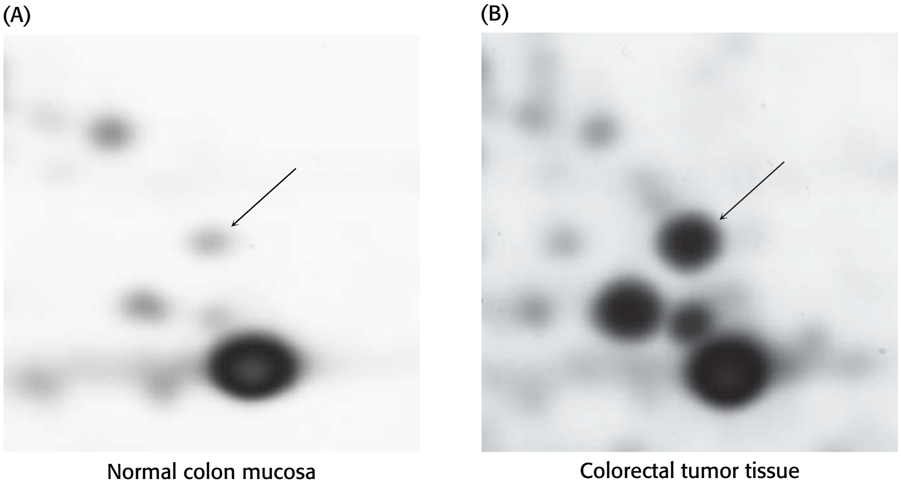


FIGURE 3.13 Alterations in protein levels detected by two-dimensional gel electrophoresis. Samples of (A) normal colon mucosa and (B) colorectal tumor tissue from the same person were analyzed by two-dimensional gel electrophoresis. In the gel section shown, changes in the intensity of several spots are evident, including a dramatic increase in levels of the protein indicated by the arrow, corresponding to the enzyme glyceraldehyde-3-phosphate dehydrogenase. [Courtesy of Lin Quinsong © 2010, The American Society for Biochemistry and Molecular Biology.]

TABLE 3.1 Quantification of a purification protocol for a fictitious protein

Step	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Gel-filtration chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000

spectrometric techniques. We will examine these powerful techniques shortly (Section 3.3).

A protein purification scheme can be quantitatively evaluated

To determine the success of a protein purification scheme, we monitor each step of the procedure by determining the specific activity of the protein mixture and by subjecting it to SDS-PAGE analysis. Consider the results for the purification of a fictitious protein, summarized in Table 3.1 and Figure 3.14. At each step, the following parameters are measured:

Total Protein. The quantity of protein present in a fraction is obtained by determining the protein concentration of a part of each fraction and multiplying by the fraction's total volume.

Total Activity. The enzyme activity for the fraction is obtained by measuring the enzyme activity in the volume of fraction used in the assay and multiplying by the fraction's total volume.

Specific Activity. This parameter is obtained by dividing total activity by total protein.

Yield. This parameter is a measure of the activity retained after each purification step as a percentage of the activity in the crude extract. The amount of activity in the initial extract is taken to be 100%.

Purification Level. This parameter is a measure of the increase in purity and is obtained by dividing the specific activity, calculated after each purification step, by the specific activity of the initial extract.

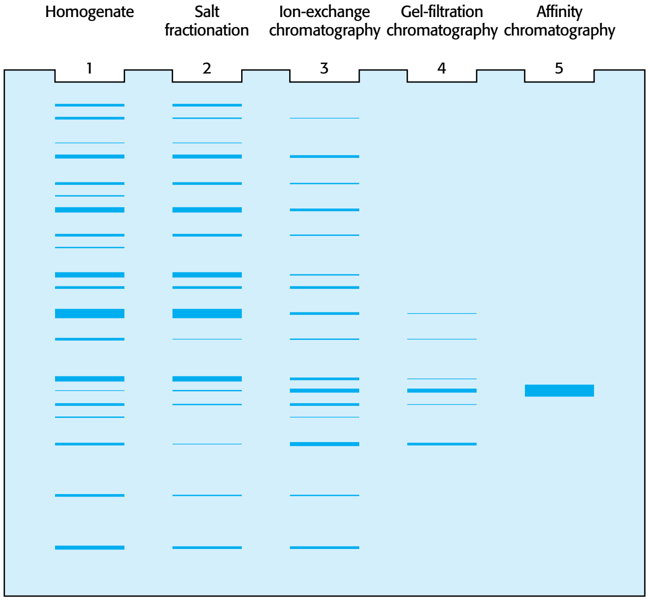


FIGURE 3.14 Electrophoretic analysis of a protein purification. The purification scheme in Table 3.1 was analyzed by SDS-PAGE. Each lane contained 50 μg of sample. The effectiveness of the purification can be seen as the band for the protein of interest becomes more prominent relative to other bands.

As we see in Table 3.1, the first purification step, salt fractionation, leads to an increase in purity of only 3-fold, but we recover nearly all the target protein in the original extract, given that the yield is 92%. After dialysis to lower the high concentration of salt remaining from the salt fractionation, the fraction is passed through an ion-exchange column. The purification now increases to 9-fold compared with the original extract, whereas the yield falls to 77%. Gel-filtration chromatography brings the level of purification to 110-fold, but the yield is now at 50%. The final step is affinity chromatography with the use of a ligand specific for the target enzyme. This step, the most powerful of these purification procedures, results in a purification level of 3000-fold but lowers the yield to 35%. The SDS-PAGE analysis in Figure 3.14 shows that, if we load a constant amount of protein onto each lane after each step, the number of bands decreases in proportion to the level of purification, and the amount of protein of interest increases as a proportion of the total protein present.

A good purification scheme takes into account both purification levels and yield. A high degree of purification and a poor yield leave little protein with which to experiment. A high yield with low purification leaves many contaminants (proteins other than the one of interest) in the fraction and complicates the interpretation of subsequent experiments.

Ultracentrifugation is valuable for separating biomolecules and determining their masses

We have already seen that centrifugation is a powerful and generally applicable method for separating a crude mixture of cell components. This technique is also valuable for the analysis of the physical properties of biomolecules. Using centrifugation, we can determine such parameters as mass and density, learn about the shape of a molecule, and investigate the interactions between molecules. To deduce these properties from the centrifugation data, we require a mathematical description of how a particle behaves when a centrifugal force is applied.

A particle will move through a liquid medium when subjected to a centrifugal force. A convenient means of quantifying the rate of movement is to calculate the sedimentation coefficient, s, of a particle by using the following equation:

$$s = m(1 - \bar{\nu}\rho)/f$$

where m is the mass of the particle, $\bar{\nu}$ is the partial specific volume (the reciprocal of the particle density), ρ is the density of the medium, and f is the frictional coefficient (a measure of the shape of the particle). The $(1 - \bar{\nu}\rho)$ term is the buoyant force exerted by liquid medium.

Sedimentation coefficients are usually expressed in *Svedberg units* (S), equal to 10^{-13} s. The smaller the S value, the more slowly a molecule moves in a centrifugal field. The S values for a number of biomolecules and cellular components are listed in Table 3.2 and Figure 3.15.

Several important conclusions can be drawn from the preceding equation:

1. The sedimentation velocity of a particle depends in part on its mass. A more massive particle sediments more rapidly than does a less massive particle of the same shape and density.

2. Shape, too, influences the sedimentation velocity because it affects the viscous drag. The frictional coefficient f of a compact particle is smaller than that of an extended particle of the same mass. Hence, elongated particles sediment more slowly than do spherical ones of the same mass.

TABLE 3.2 S values and molecular weights of sample proteins

Protein	S value (Svedberg units)	Molecular weight
Pancreatic trypsin inhibitor	1	6,520
Cytochrome *c*	1.83	12,310
Ribonuclease A	1.78	13,690
Myoglobin	1.97	17,800
Trypsin	2.5	23,200
Carbonic anhydrase	3.23	28,800
Concanavalin A	3.8	51,260
Malate dehydrogenase	5.76	74,900
Lactate dehydrogenase	7.54	146,200

Source: T. Creighton, *Proteins,* 2d ed. (W. H. Freeman and Company, 1993), Table 7.1.

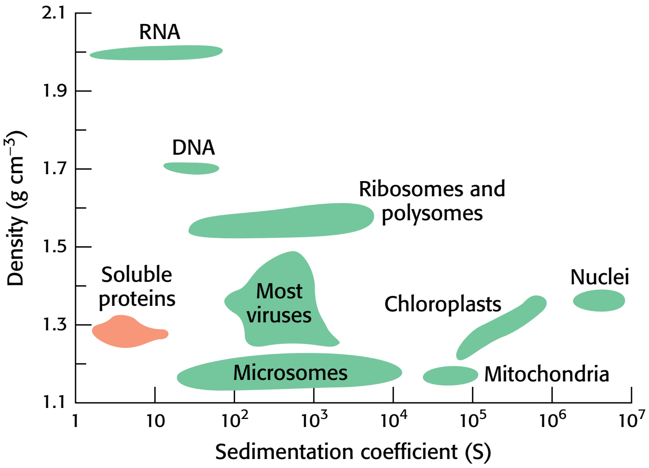


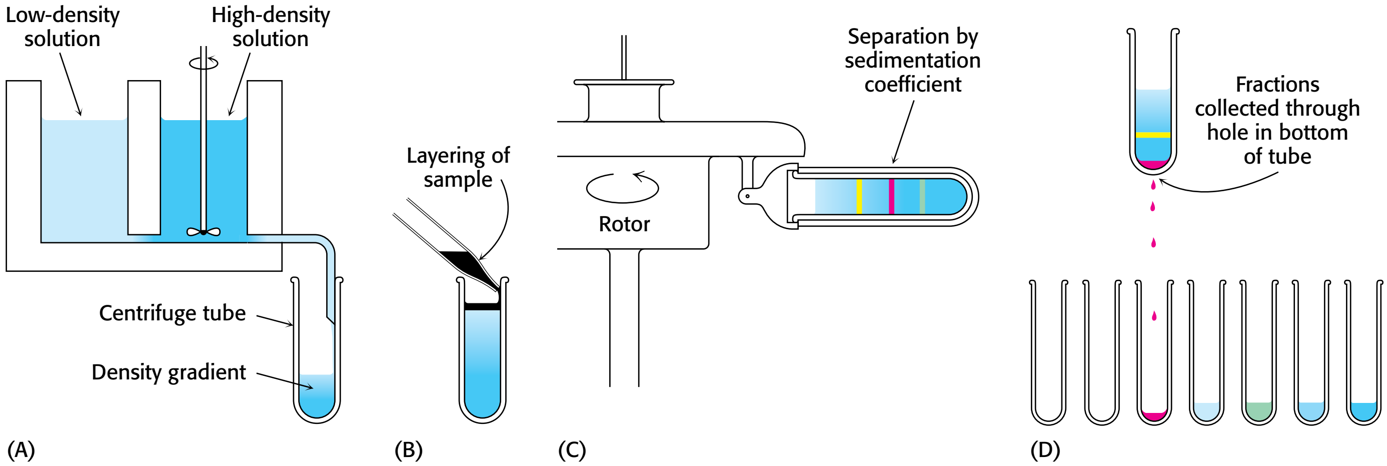
FIGURE 3.15 Density and sedimentation coefficients of cellular components. [Data from L. J. Kleinsmith and V. M. Kish, *Principles of Cell and Molecular Biology,* 2d ed. (HarperCollins, 1995), p. 138.]

3. A dense particle moves more rapidly than does a less dense one because the opposing buoyant force $(1 - \bar{\nu}\rho)$ is smaller for the denser particle.

4. The sedimentation velocity also depends on the density of the solution (ρ). Particles sink when $\bar{\nu}\rho < 1$, float when $\bar{\nu}\rho > 1$, and do not move when $\bar{\nu}\rho = 1$.

A technique called *zonal, band,* or most commonly *gradient* centrifugation can be used to separate proteins with different sedimentation coefficients. The first step is to form a density gradient in a centrifuge tube. Differing proportions of a low-density solution (such as 5% sucrose) and a high-density solution (such as 20% sucrose) are mixed to create a linear gradient of sucrose concentration ranging from 20% at the bottom of the tube to 5% at the top (Figure 3.16). The role of the gradient is to prevent convective flow. A small volume of a solution containing the mixture of proteins to be separated is placed on top of the density gradient. When the rotor is spun, proteins move through the gradient and separate according to their sedimentation coefficients. The time and speed of the centrifugation is determined empirically. The separated bands, or zones, of protein can be harvested by making a hole in the bottom of the tube and collecting drops. The drops can be measured for protein content and catalytic activity or another functional property. This sedimentation-velocity technique readily separates proteins differing in sedimentation coefficient by a factor of two or more.

FIGURE 3.16 Zonal centrifugation. The steps are as follows: (A) form a density gradient, (B) layer the sample on top of the gradient, (C) place the tube in a swinging-bucket rotor and centrifuge it, and (D) collect the samples. [Information from D. Freifelder, *Physical Biochemistry*, 2d ed. (W. H. Freeman and Company, 1982), p. 397.]



The mass of a protein can be directly determined by *sedimentation equilibrium,* in which a sample is centrifuged at low speed such that a concentration gradient of the sample is formed. However, this sedimentation is counterbalanced by the diffusion of the sample from regions of high to low concentration. When equilibrium has been achieved, the shape of the final gradient depends solely on the mass of the sample. *The sedimentation-equilibrium technique for determining mass is very accurate and can be applied without denaturing the protein. Thus the native quaternary structure of multimeric proteins is preserved.* In contrast, SDS–polyacrylamide gel electrophoresis provides an *estimate* of the mass of dissociated polypeptide chains under *denaturing* conditions. Note that, if we know the mass of the dissociated components of a multimeric protein as determined by SDS–polyacrylamide analysis and the mass of the intact multimer as determined by sedimentation-equilibrium analysis, we can determine the number of copies of each polypeptide chain present in the protein complex.

Protein purification can be made easier with the use of recombinant DNA technology

In Chapter 5, we shall consider the widespread effect of recombinant DNA technology on all areas of biochemistry and molecular biology. The application of recombinant methods to the overproduction of proteins has enabled dramatic advances in our understanding of their structure and function. Before the advent of this technology, proteins were isolated solely from their native sources, often requiring a large amount of tissue to obtain a sufficient amount of protein for analytical study. For example, the purification of bovine deoxyribonuclease in 1946 required nearly ten pounds of beef pancreas to yield one gram of protein. As a result, biochemical studies on purified material were often limited to abundant proteins.

Armed with the tools of recombinant technology, the biochemist is now able to enjoy a number of significant advantages:

1. *Proteins can be expressed in large quantities.* The homogenate serves as the starting point in a protein purification scheme. For recombinant systems, a host organism that is amenable to genetic manipulation, such as the bacterium *Escherichia coli* or the yeast *Pichia pastoris,* is utilized to express a protein of interest. The biochemist can exploit the short doubling times and ease of genetic manipulation of such organisms to produce large amounts of protein from manageable amounts of culture. As a result, purification can begin with a homogenate that is often highly enriched with the desired molecule. Moreover, a protein can be easily obtained regardless of its natural abundance or its species of origin.

2. *Affinity tags can be fused to proteins.* As described above, affinity chromatography can be a highly selective step within a protein purification scheme. Recombinant DNA technology enables the attachment of any one of a number of possible affinity tags to a protein (such as the "His tag" mentioned earlier). Hence, the benefits of affinity chromatography can be realized even for those proteins for which a binding partner is unknown or not easily determined.

3. *Proteins with modified primary structures can be readily generated.* A powerful aspect of recombinant DNA technology as applied to protein purification is the ability to manipulate genes to generate variants of a native protein sequence (Section 5.2). We learned in Section 2.4 that many proteins consist of compact domains connected by flexible linker regions. With the use of genetic-manipulation strategies, fragments of a protein that encompass single domains can be generated, an advantageous approach when expression of the

entire protein is limited by its size or solubility. Additionally, as we will see in Section 9.1, amino acid substitutions can be introduced into the active site of an enzyme to precisely probe the roles of specific residues within its catalytic cycle.

3.2 Immunology Provides Important Techniques with Which to Investigate Proteins

The purification of a protein enables the biochemist to explore its function and structure within a precisely controlled environment. However, the isolation of a protein removes it from its native context within the cell, where its activity is most physiologically relevant. Advances in the field of immunology (Chapter 34) have enabled the use of antibodies as critical reagents for exploring the functions of proteins within the cell. The exquisite specificity of antibodies for their target proteins provides a means to tag a specific protein so that it can be isolated, quantified, or visualized.

Antibodies to specific proteins can be generated

Immunological techniques begin with the generation of antibodies to a particular protein. An *antibody* (also called an *immunoglobulin,* Ig) is itself a protein (Figure 3.17); it is synthesized by vertebrates in response to the presence of a foreign substance, called an *antigen*. Antibodies have specific and high affinity for the antigens that elicited their synthesis. The binding of antibody to antigen is a step in the immune response that protects the animal from infection (Chapter 34). Foreign proteins, polysaccharides, and nucleic acids can be antigens. Small foreign molecules, such as synthetic peptides, also can elicit antibodies, provided that the small molecule is attached to a macromolecular carrier. An antibody recognizes a specific group or cluster of amino acids on the target molecule called an *antigenic determinant* or *epitope*. The specificity of the antibody–antigen interaction is a consequence of the shape complementarity between the two surfaces (Figure 3.18). Animals have a very large repertoire of antibody-producing cells, each producing an antibody that contains a unique surface for antigen recognition. When an antigen is introduced into an animal, it is recognized by a select few cells from this population, stimulating the proliferation of these cells. This process ensures that more antibodies of the appropriate specificity are produced.

FIGURE 3.17 Antibody structure. (A) Immunoglobulin G (IgG) consists of four chains, two heavy chains (blue) and two light chains (red), linked by disulfide bonds. The heavy and light chains come together to form F_{ab} domains, which have the antigen-binding sites at the ends. The two heavy chains form the F_c domain. *Notice* that the F_{ab} domains are linked to the F_c domain by flexible linkers. (B) A more schematic representation of an IgG molecule. [Drawn from 1IGT.pdb.]

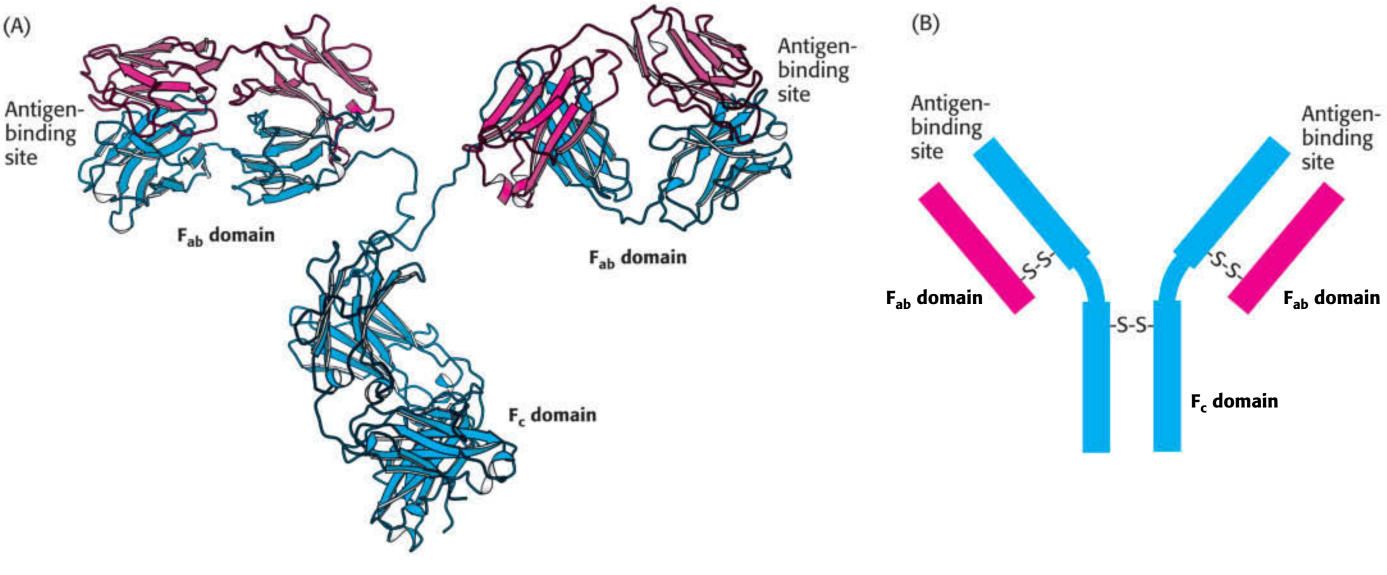
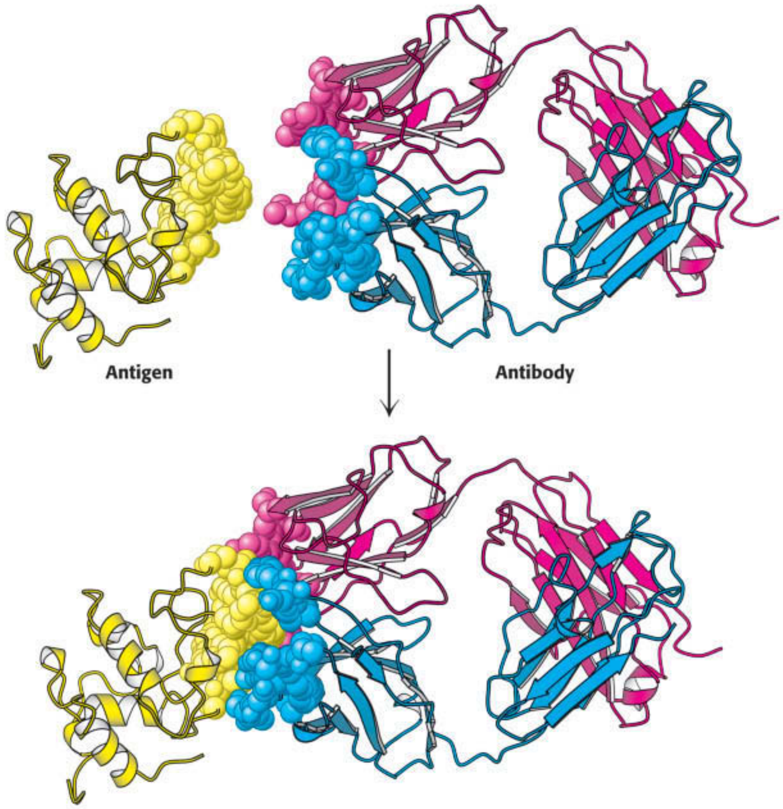


FIGURE 3.18 Antigen–antibody interactions. A protein antigen, in this case lysozyme, binds to the end of an F_{ab} domain of an antibody. *Notice* that the end of the antibody and the antigen have complementary shapes, allowing a large amount of surface to be buried on binding. [Drawn from 1YQV.pdb.]



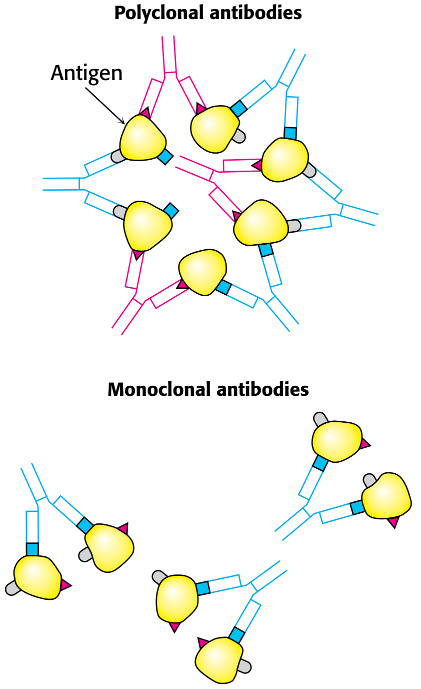


FIGURE 3.19 Polyclonal and monoclonal antibodies. Most antigens have several epitopes. Polyclonal antibodies are heterogeneous mixtures of antibodies, each specific for one of the various epitopes on an antigen. Monoclonal antibodies are all identical, produced by clones of a single antibody-producing cell. They recognize one specific epitope. [Information from R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology,* 4th ed. (W. H. Freeman and Company, 2000), p. 154.]

Immunological techniques depend on the ability to generate antibodies to a specific antigen. To obtain antibodies that recognize a particular protein, a biochemist injects the protein into a rabbit twice, 3 weeks apart. The injected protein acts as an antigen, stimulating the reproduction of cells producing antibodies that recognize it. Blood is drawn from the immunized rabbit several weeks later and centrifuged to separate blood cells from the supernatant, or serum. The serum, called an *antiserum,* contains antibodies to all antigens to which the rabbit has been exposed. Only some of them will be antibodies to the injected protein. Moreover, antibodies that recognize a particular antigen are not a single molecular species. For instance, 2,4-dinitrophenol (DNP) was used as an antigen to generate antibodies. Analyses of anti-DNP antibodies revealed a wide range of binding affinities; the dissociation constants ranged from about 0.1 nM to 1 μM. Correspondingly, a large number of bands were evident when anti-DNP antibody was subjected to isoelectric focusing. These results indicate that cells are producing many different antibodies, each recognizing a different surface feature of the same antigen. These antibodies are termed *polyclonal,* referring to the fact that they are derived from multiple antibody-producing cell populations (Figure 3.19). The heterogeneity of polyclonal antibodies can be advantageous for certain applications, such as the detection of a protein of low abundance, because each protein molecule can be bound by more than one antibody at multiple distinct antigenic sites.

Monoclonal antibodies with virtually any desired specificity can be readily prepared

The discovery of a means of producing *monoclonal antibodies* of virtually any desired specificity was a major breakthrough that intensified the power of immunological approaches. As with impure proteins, working with an impure mixture of antibodies makes it difficult to interpret data. Ideally, one would isolate a clone of cells producing a single, identical antibody. The problem is that antibody-producing cells isolated from an organism have short life spans.

Immortal cell lines that produce monoclonal antibodies do exist. These cell lines are derived from a type of cancer, *multiple myeloma*, which is a malignant disorder of antibody-producing cells. In this cancer, a single transformed plasma cell divides uncontrollably, generating a very large number of *cells of a single kind*. Such a group of cells is a *clone* because the cells are descended from the same cell and have identical properties. The identical cells of the myeloma secrete large amounts of *a single immunoglobulin* generation after generation. While these antibodies have proven useful for elucidating antibody structure, nothing is known about their specificity. Hence, they have little utility for the immunological methods described in the next pages.

César Milstein and Georges Köhler discovered that large amounts of antibodies of nearly any desired specificity can be obtained by fusing a short-lived antibody-producing cell with an immortal myeloma cell. An antigen is injected into a mouse, and its spleen is removed several weeks later (Figure 3.20). A mixture of plasma cells from this spleen is fused in vitro with myeloma cells. Each of the resulting hybrid cells, called hybridoma cells, indefinitely produces the identical antibody specified by the parent cell from the spleen. Hybridoma cells can then be screened by a specific assay for the antigen–antibody interaction to determine which ones produce antibodies of the preferred specificity. Collections of cells shown to produce the desired antibody are subdivided and reassayed. This process is repeated until a pure cell line, a clone producing a single antibody, is isolated. These positive cells can be grown in

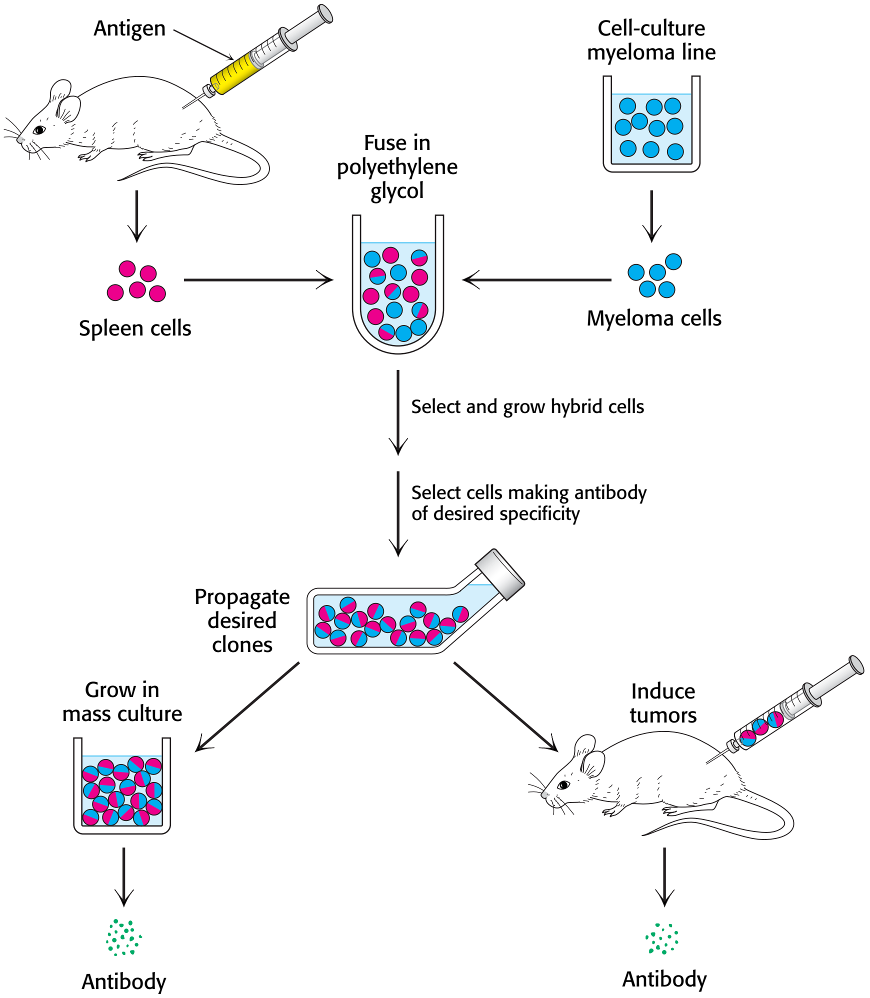

Antigen
Cell-culture myeloma line
Fuse in polyethylene glycol
Spleen cells
Myeloma cells
Select and grow hybrid cells
Select cells making antibody of desired specificity
Propagate desired clones
Grow in mass culture
Induce tumors
Antibody
Antibody

FIGURE 3.20 Preparation of monoclonal antibodies. Hybridoma cells are formed by the fusion of antibody-producing cells and myeloma cells. The hybrid cells are allowed to proliferate by growing them in selective medium. They are then screened to determine which ones produce antibody of the desired specificity. [Information from C. Milstein. Monoclonal antibodies. Copyright © 1980 by Scientific American, Inc. All rights reserved.]

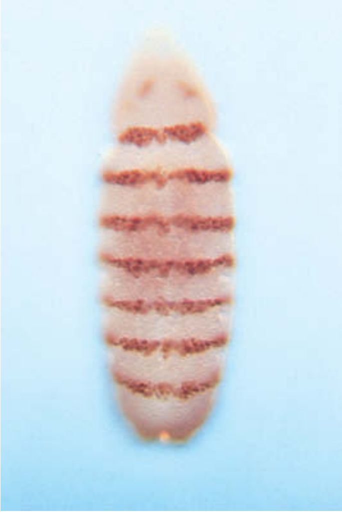


FIGURE 3.21 Fluorescence micrograph of a developing *Drosophila* embryo. The embryo was stained with a fluorescence-labeled monoclonal antibody for the DNA-binding protein encoded by *engrailed,* an essential gene in specifying the body plan. [Courtesy of Dr. Nipam Patel and Dr. Corey Goodman.]

culture medium or injected into mice to induce myelomas. Alternatively, the cells can be frozen and stored for long periods.

The hybridoma method of producing monoclonal antibodies has opened new vistas in biology and medicine. Large amounts of identical antibodies with tailor-made specificities can be readily prepared. They are sources of insight into relations between antibody structure and specificity. Moreover, monoclonal antibodies can serve as precise analytical and preparative reagents. Proteins that guide development have been identified with the use of monoclonal antibodies as tags (Figure 3.21). Monoclonal antibodies attached to solid supports can be used as affinity columns to purify scarce proteins. This method has been used to purify interferon (an antiviral protein) 5000-fold from a crude mixture. Clinical laboratories use monoclonal antibodies in many assays. For example, the detection in blood of isozymes that are normally localized in the heart points to a myocardial infarction (heart attack). Blood transfusions have been made safer by antibody screening of donor blood for viruses that cause AIDS (acquired immune deficiency syndrome), hepatitis, and other infectious diseases. Monoclonal antibodies can be used as therapeutic agents. For example, trastuzumab (Herceptin) is a monoclonal antibody useful for treating some forms of breast cancer.

Proteins can be detected and quantified by using an enzyme-linked immunosorbent assay

Antibodies can be used as exquisitely specific analytic reagents to quantify the amount of a protein or other antigen present in a biological sample. The *enzyme-linked immunosorbent assay* (ELISA) makes use of an enzyme that reacts with a colorless substrate to produce a colored product. The enzyme is covalently linked to a specific antibody that recognizes a target antigen. If the antigen is present, the antibody–enzyme complex will bind to it and, on addition of the substrate, the enzyme will catalyze the reaction, generating the colored product. Thus, the presence of the colored product indicates the presence of the antigen. Rapid and convenient, ELISAs can detect less than a nanogram (10^{-9} g) of a specific protein. ELISAs can be performed with either polyclonal or monoclonal antibodies, but the use of monoclonal antibodies yields more-reliable results.

We will consider two among the several types of ELISA. *The indirect ELISA is used to detect the presence of antibody* and is the basis of the test for HIV infection. The HIV test detects the presence of antibodies that recognize viral core protein antigens. Viral core proteins are adsorbed to the bottom of a well. Antibodies from the person being tested are then added to the coated well. Only someone infected with HIV will have antibodies that bind to the antigen. Finally, enzyme-linked antibodies to human antibodies (e.g., enzyme-linked goat antibodies that recognize human antibodies) are allowed to react in the well, and unbound antibodies are removed by washing. Substrate is then applied. An enzyme reaction yielding a colored product suggests that the enzyme-linked antibodies were bound to human antibodies, which in turn implies that the patient has antibodies to the viral antigen (Figure 3.22A). This assay is quantitative: the rate of the color-formation reaction is proportional to the amount of antibody originally present.

The sandwich ELISA is used to detect antigen rather than antibody. Antibody to a particular antigen is first adsorbed to the bottom of a well. Next, solution containing the antigen (such as blood or urine, in medical diagnostic tests) is added to the well and binds to the antibody. Finally, a

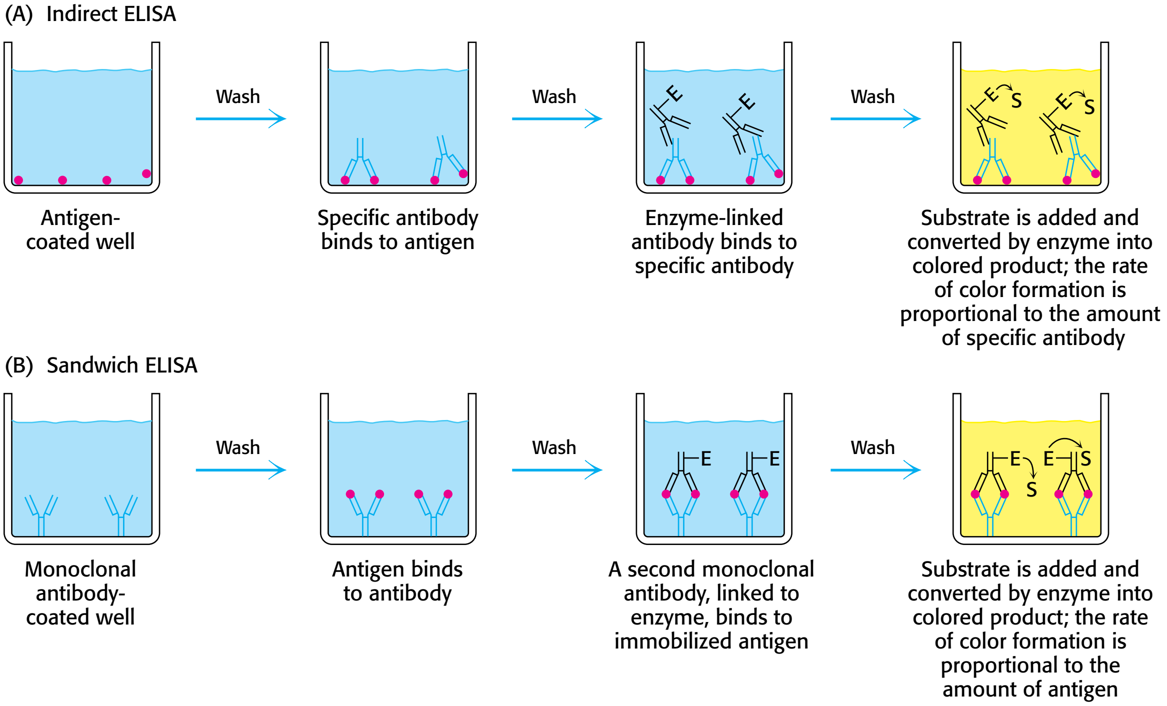

(A) Indirect ELISA
Wash
Wash
Wash
E
E
E
S
E
S
Antigen-coated well
Specific antibody binds to antigen
Enzyme-linked antibody binds to specific antibody
Substrate is added and converted by enzyme into colored product; the rate of color formation is proportional to the amount of specific antibody
(B) Sandwich ELISA
Wash
Wash
Wash
E
E
E
S
E
S
Monoclonal antibody-coated well
Antigen binds to antibody
A second monoclonal antibody, linked to enzyme, binds to immobilized antigen
Substrate is added and converted by enzyme into colored product; the rate of color formation is proportional to the amount of antigen

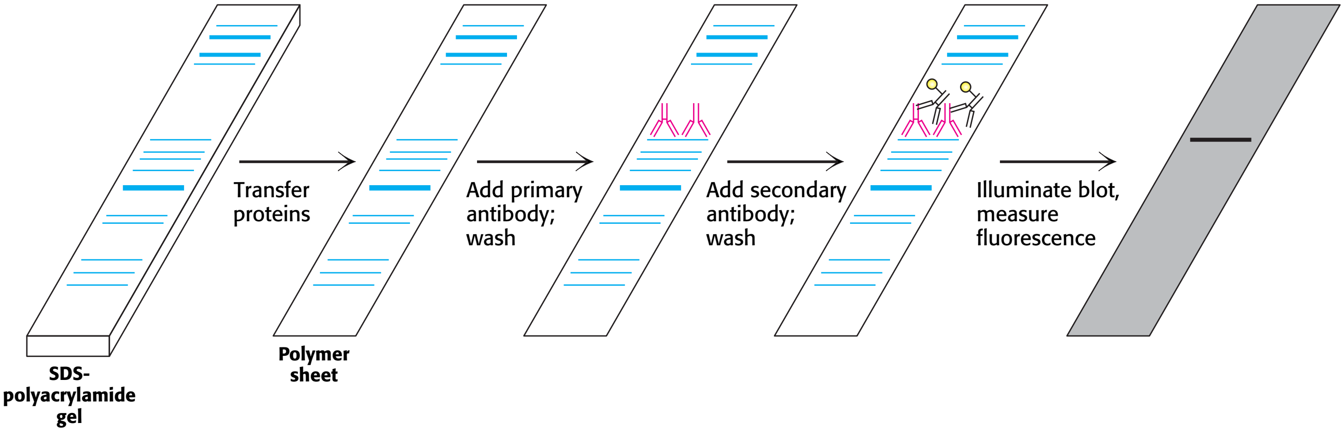
FIGURE 3.22 Indirect ELISA and sandwich ELISA. (A) In indirect ELISA, the production of color indicates the amount of an antibody to a specific antigen. (B) In sandwich ELISA, the production of color indicates the quantity of antigen. [Information from R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology,* 4th ed. (W. H. Freeman and Company, 2000), p. 162.]

second, different antibody to the antigen is added. This antibody is enzyme linked and is processed as described for indirect ELISA. In this case, the rate of color formation is directly proportional to the amount of antigen present. Consequently, it permits the measurement of small quantities of antigen (Figure 3.22B).

Western blotting permits the detection of proteins separated by gel electrophoresis

Very small quantities of a protein of interest in a cell or in body fluid can be detected by an immunoassay technique called *western blotting* (Figure 3.23). A sample is subjected to electrophoresis on an SDS–polyacrylamide gel. A polymer sheet is pressed against the gel, transferring the resolved proteins on the gel to the sheet, which makes the proteins more accessible for reaction. An antibody that is specific for the protein of interest, called the *primary*

FIGURE 3.23 Western blotting. Proteins on an SDS–polyacrylamide gel are transferred to a polymer sheet. The sheet is first treated with a primary antibody, which is specific for the protein of interest, and then washed to remove unbound antibody. Next, the sheet is treated with a secondary antibody, which recognizes the primary antibody, and washed again. Since the secondary antibody is labeled (here, with a fluorescent tag indicated by the yellow circle), the band containing the protein of interest can be identified.


Transfer proteins
Add primary antibody; wash
Add secondary antibody; wash
Illuminate blot, measure fluorescence
SDS-polyacrylamide gel
Polymer sheet

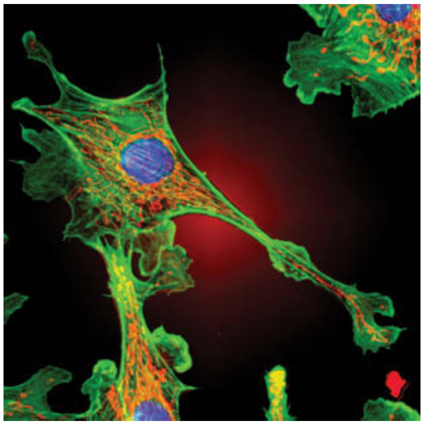


FIGURE 3.24 **Actin filaments.** Fluorescence micrograph of a cell shows actin filaments stained green using an antibody specific to actin. [David Becker/Science Source.]

antibody, is added to the sheet and reacts with the antigen. The antibody–antigen complex on the sheet can then be detected by rinsing the sheet with a second antibody, called the *secondary antibody*, that is specific for the primary antibody (e.g., a goat antibody that recognizes mouse antibodies). Typically, the secondary antibody is fused to an enzyme that produces a chemiluminescent or colored product or contains a fluorescent tag, enabling the identification and quantitation of the protein of interest. Western blotting makes it possible to find a protein in a complex mixture, the proverbial needle in a haystack. It is the basis for the test for infection by hepatitis C, where it is used to detect a core protein of the virus. This technique is also very useful in monitoring protein purification and in the cloning of genes.

Fluorescent markers make the visualization of proteins in the cell possible

Biochemistry is often performed in test tubes or polyacrylamide gels. However, most proteins function in the context of a cell. Fluorescent markers provide a powerful means of examining proteins in their biological context. Cells can be stained with fluorescence-labeled antibodies and examined by *fluorescence microscopy* to reveal the location of a protein of interest. For example, arrays of parallel bundles are evident in cells stained with antibody specific for actin, a protein that polymerizes into filaments (Figure 3.24). Actin filaments are constituents of the cytoskeleton, the internal scaffolding of cells that controls their shape and movement. By tracking protein location, fluorescent markers also provide clues to protein function. For instance, the mineralocorticoid receptor protein binds to steroid hormones (Section 26.4), including cortisol. The receptor was linked to a yellow variant of *green fluorescent protein* (GFP), a naturally fluorescent protein isolated from the jellyfish *Aequorea victoria* (Chapter 2). Fluorescence microscopy revealed that, in the absence of the hormone, the receptor is located in the cytoplasm (Figure 3.25A). On addition of the steroid, the receptor is translocated to the nucleus, where it binds to DNA (Figure 3.25B). These results indicate that the mineralocorticoid receptor protein is a transcription factor that controls gene expression.

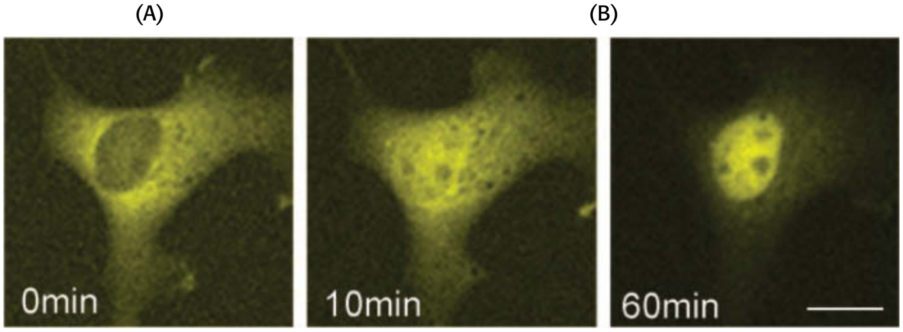


FIGURE 3.25 **Nuclear localization of a steroid receptor.** (A) The mineralocorticoid receptor, made visible by attachment to a yellow variant of GFP, is located predominantly in the cytoplasm of the cultured cell. (B) Subsequent to the addition of corticosterone (a glucocorticoid steroid that also binds to the mineralocorticoid receptor), the receptor moves into the nucleus. [M. Nishi, M. Tanaka, K.-i. Matsuda, M. Sunaguchi, and M. Kawata. *J. Neurosci.* 24:4918–4927, 2004, Fig. 7A.]

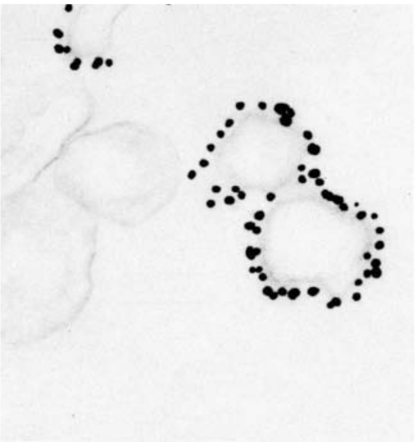


FIGURE 3.26 **Immunoelectron microscopy.** The opaque particles (150-Å, or 15-nm, diameter) in this electron micrograph are clusters of gold atoms bound to antibody molecules. A gold-labeled antibody against a channel protein (Section 13.4) identifies membrane vesicles at the termini of neurons that contain this protein. [Courtesy of Dr. Peter Sargent.]

The highest resolution of fluorescence microscopy is about 0.2 μm (200 nm, or 2000 Å), the wavelength of visible light. Finer spatial resolution can be achieved by electron microscopy if the antibodies are tagged with electron-dense markers. For example, antibodies conjugated to clusters of gold or to ferritin (which has an electron-dense core rich in iron) are highly visible under the electron microscope. *Immunoelectron microscopy* can define the position of antigens to a resolution of 10 nm (100 Å) or finer (Figure 3.26).

3.3 Mass Spectrometry is a Powerful Technique for the Identification of Peptides and Proteins

In many instances, the study of a particular biological process in its native context is advantageous. For example, if we are interested in a pathway that is localized to the nucleus of a cell, we might conduct studies on an isolated nuclear extract. In these experiments, identification of the proteins present in the sample is often critical. Antibody-based techniques, such as the ELISA method described in the previous section, can be very helpful toward this goal. However, these techniques are limited to the detection of proteins for which an antibody is already available. Mass spectrometry enables the highly precise and sensitive measurement of the atomic composition of a particular molecule, or *analyte,* without prior knowledge of its identity. Originally, this method was relegated to the study of the chemical composition and molecular mass of gases or volatile liquids. However, technological advances in the past two decades have dramatically expanded the utility of mass spectrometry to the study of proteins, even those found at very low concentrations within highly complex mixtures, such as the contents of a particular cell type.

Mass spectrometry enables the highly accurate and sensitive detection of the mass of an analyte. This information can be used to determine the identity and chemical state of the molecule of interest. Mass spectrometers operate by converting analyte molecules into gaseous, charged forms (*gas-phase ions*). Through the application of electrostatic potentials, the ratio of the mass of each ion to its charge (the *mass-to-charge ratio*, or m/z) can be measured. Although a wide variety of techniques employed by mass spectrometers are used in current practice, each of them comprises three essential components: the ion source, the mass analyzer, and the detector. Let us consider the first two in greater detail, because improvements in them have contributed most significantly to the analysis of biological samples.

The *ion source* achieves the first critical step in mass spectrometric analysis: conversion of the analyte into gas-phase ions (*ionization*). Until recently, proteins could not be ionized efficiently because of their high molecular weights and low volatility. However, the development of techniques such as *matrix-assisted laser desorption/ionization* (MALDI) and *electrospray ionization* (ESI) has enabled the clearing of this significant hurdle. In MALDI, the analyte is evaporated to dryness in the presence of a volatile, aromatic compound (the *matrix*) that can absorb light at specific wavelengths. A laser pulse tuned to one of these wavelengths excites and vaporizes the matrix, converting some of the analyte into the gas phase. Subsequent gaseous collisions enable the intermolecular transfer of charge, ionizing the analyte. In ESI, a solution of the analyte is passed through an electrically charged nozzle. Droplets of the analyte, now charged, emerge from the nozzle into a chamber of very low pressure, evaporating the solvent and ultimately yielding the ionized analyte.

The newly formed analyte ions then enter the *mass analyzer,* where they are distinguished on the basis of their mass-to-charge ratios. There are a number of different types of mass analyzers. For this discussion, we will consider one of the simplest, the *time-of-flight* (TOF) *mass analyzer*, in which ions are accelerated through an elongated chamber under a fixed electrostatic potential. Given two ions of identical net charge, the smaller ion will require less time to traverse the chamber than will the larger ion. The mass of each ion can be determined by measuring the time required for each ion to pass through the chamber.

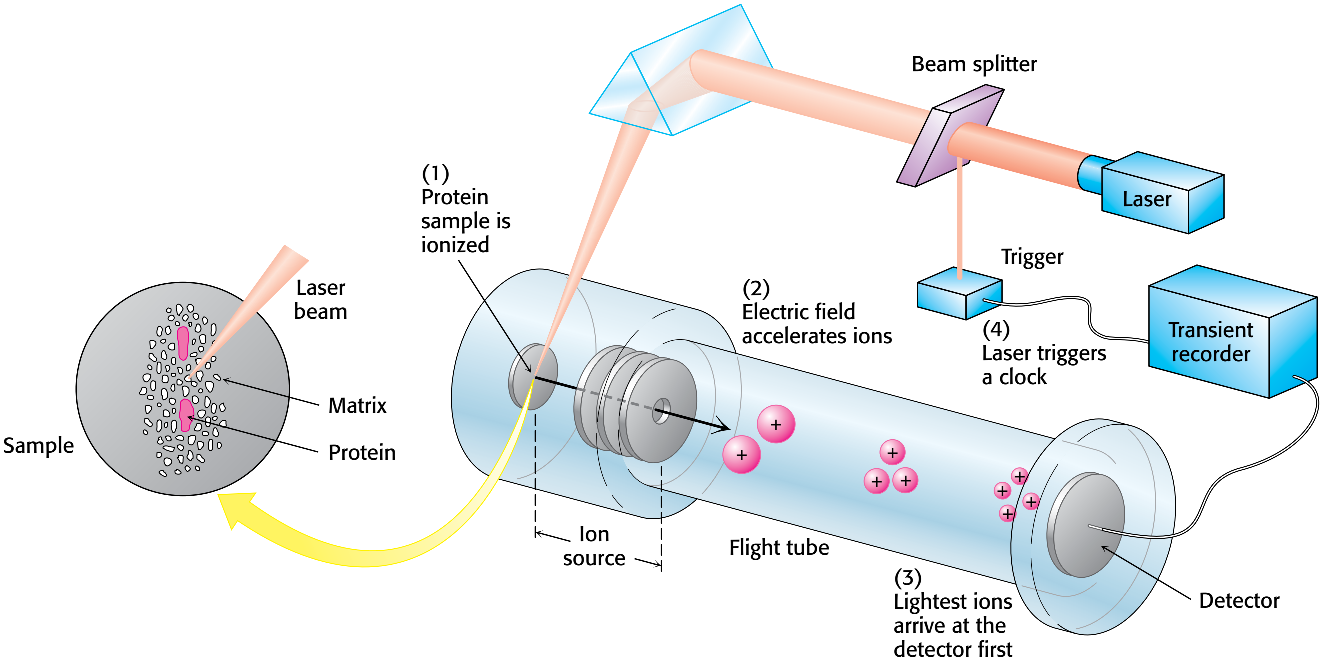
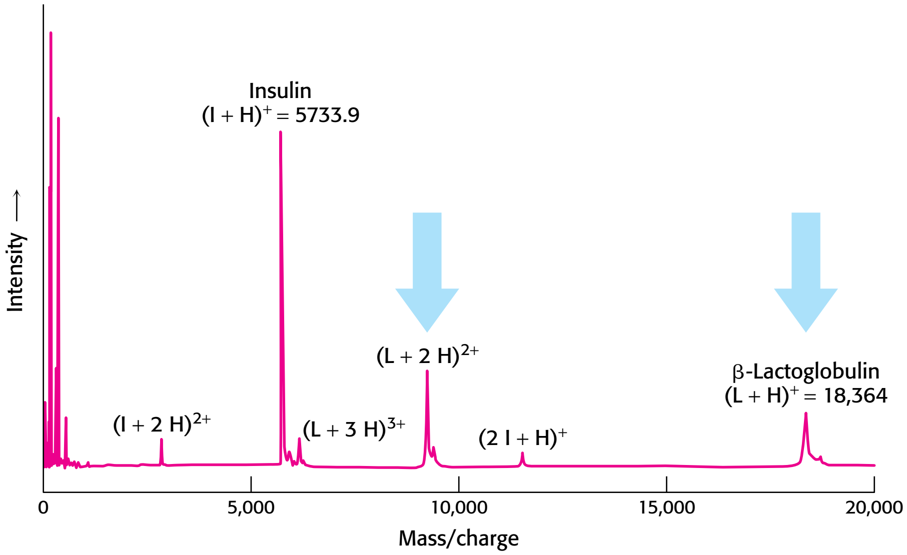

Beam splitter
Laser
(1) Protein sample is ionized
Laser beam
Trigger
(2) Electric field accelerates ions
(4) Laser triggers a clock
Transient recorder
Sample
Matrix
Protein
Ion source
Flight tube
(3) Lightest ions arrive at the detector first
Detector

FIGURE 3.27 MALDI-TOF mass spectrometry. (1) The protein sample, embedded in an appropriate matrix, is ionized by the application of a laser beam. (2) An electric field accelerates the ions through the flight tube toward the detector. (3) The lightest ions arrive first. (4) The ionizing laser pulse also triggers a clock that measures the time of flight (TOF) for the ions. [Information from J. T. Watson, *Introduction to Mass Spectrometry,* 3d ed. (Lippincott-Raven, 1997), p. 279.]

The sequential action of the ion source and the mass analyzer enables the highly sensitive measurement of the mass of potentially massive ions, such as those of proteins. Consider an example of a MALDI ion source coupled to a TOF mass analyzer: the MALDI-TOF mass spectrometer (Figure 3.27). Gas-phase ions generated by the MALDI ion source pass directly into the TOF analyzer, where the mass-to-charge ratios are recorded. In Figure 3.28, the MALDI-TOF mass spectrum of a mixture of 5 pmol each of insulin and lactoglobulin is shown. The masses determined by MALDI-TOF are 5733.9 and 18,364, respectively. A comparison with the calculated values of 5733.5 and 18,388 reveals that MALDI-TOF is clearly an accurate means of determining protein mass.

In the ionization process, a family of ions, each of the same mass but carrying different total net charges, is formed from a single analyte. Because the mass spectrometer detects ions on the basis of their mass-to-charge ratio, these ions will appear as separate peaks in the mass spectrum. For example, in the mass spectrum of β-lactoglobulin shown in Figure 3.28, peaks near $m/z = 18{,}364$ (corresponding to the +1 charged ion) and $m/z = 9{,}183$

FIGURE 3.28 MALDI-TOF mass spectrum of insulin and *β*-lactoglobulin. A mixture of 5 pmol each of insulin (I) and β-lactoglobulin (L) was ionized by MALDI, which produces predominantly singly charged molecular ions from peptides and proteins—the insulin ion $(I+H)^+$ and the lactoglobulin ion $(L+H)^+$. Molecules with multiple charges, such as those for β-lactoglobulin indicated by the blue arrows, as well as small quantities of a singly charged dimer of insulin $(2\,I+H)^+$ also are produced. [Data from J. T. Watson, *Introduction to Mass Spectrometry,* 3d ed. (Lippincott-Raven, 1997), p. 282.]



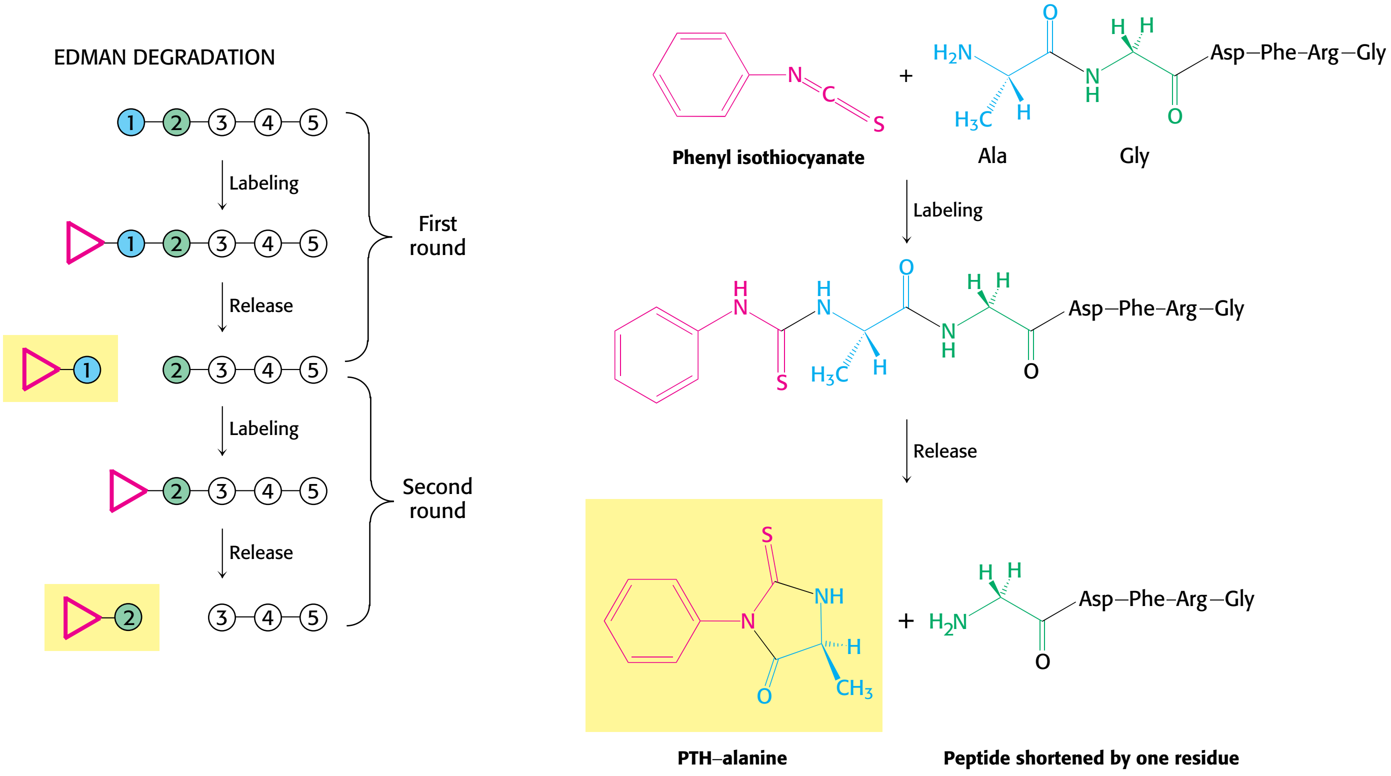
(corresponding to the +2 charged ion) are visible (indicated by the blue arrows). Although multiple peaks for the same ion may appear to be a nuisance, they enable the spectrometrist to measure the mass of an analyte ion more than once in a single experiment, improving the overall precision of the calculated result.

Peptides can be sequenced by mass spectrometry

Mass spectrometry is one of several techniques for the determination of the amino acid sequence of a protein. As we shall discuss shortly, this sequence data can be a valuable source of information. For many years, chemical methods were the primary means for peptide sequencing. In the most common of these methods, *Edman degradation*, the N-terminal amino acid of a polypeptide is labeled with *phenyl isothiocyanate*. Subsequent cleavage yields the phenylthiohydantoin (PTH)-amino acid derivative, which can be identified by spectroscopic methods, and the polypeptide chain, now shortened by one residue (Figure 3.29). This procedure can then be repeated on the shortened peptide, yielding another PTH–amino acid, which can again be identified by chromatography.

While technological advancements have improved the speed and sensitivity of the Edman degradation, these parameters have largely been surpassed by the application of mass spectrometric methods. The use of mass spectrometry for protein sequencing takes advantage of the fact that ions of proteins that have been analyzed by a mass spectrometer, the *precursor ions*, can be broken into smaller peptide chains by bombardment with atoms of an inert gas such as helium or argon. These new fragments, or *product ions*, can be passed through a second mass analyzer for further mass characterization. The utilization of two mass analyzers arranged in this manner is referred to as *tandem mass spectrometry*. Importantly, the product-ion fragments are formed in chemically predictable ways that can provide clues to the amino acid sequence of the precursor ion. For polypeptide analytes, disruption of individual peptide bonds will yield two smaller peptide ions, containing the

FIGURE 3.29 The Edman degradation. The labeled amino-terminal residue (PTH–alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide.



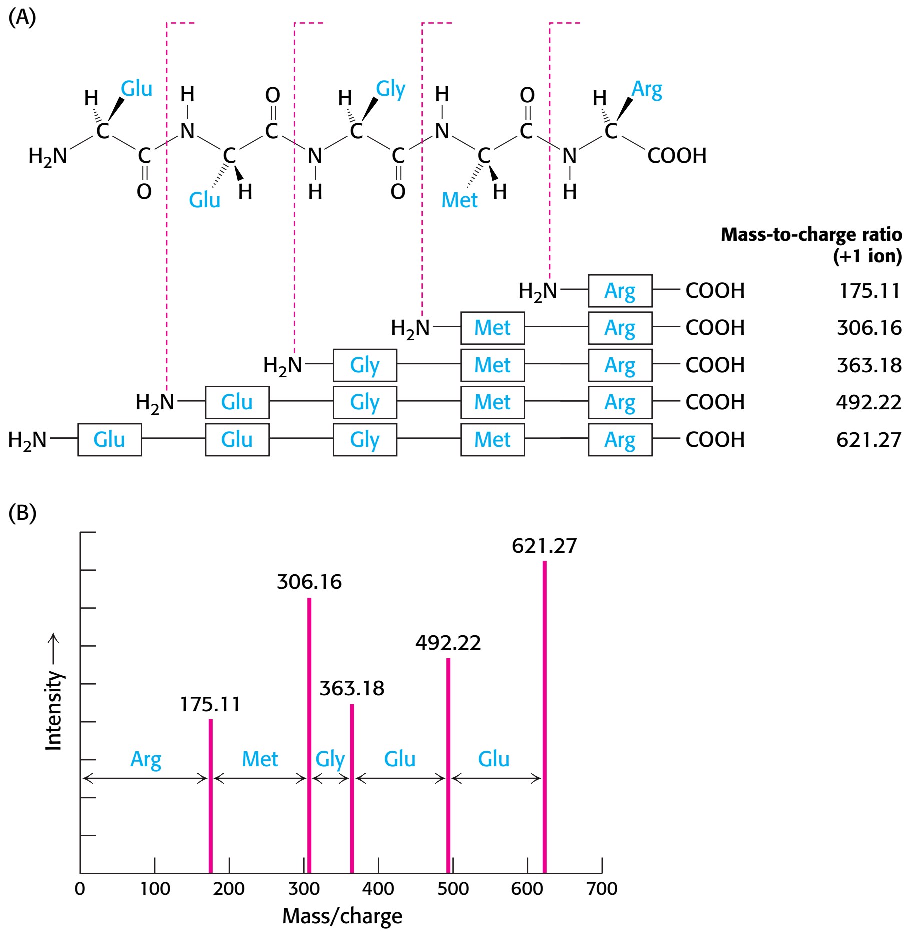


FIGURE 3.30 Peptide sequencing by tandem mass spectrometry. (A) Within the mass spectrometer, peptides can be fragmented by bombardment with inert gaseous ions to generate a family of product ions in which individual amino acids have been removed from one end. In this example, the carboxyl fragment of the cleaved peptide bond is ionized. (B) The product ions are detected in the second mass analyzer. The mass differences between the peaks indicate the sequence of amino acids in the precursor ion. [Data from H. Steen and M. Mann. *Nat. Rev. Mol. Cell Biol.* 5:699–711, 2004.]

sequences before and after the cleavage site. Hence, a family of ions can be detected; each ion represents a fragment of the original peptide with one or more amino acids removed from one end (Figure 3.30A). For simplicity, only the carboxyl-terminal peptide fragments are shown in Figure 3.30A. Figure 3.30B depicts a representative mass spectrum from a fragmented peptide. The mass differences between the peaks in this fragmentation experiment indicate the amino acid sequence of the precursor peptide ion.

Proteins can be specifically cleaved into small peptides to facilitate analysis

In principle, it should be possible to sequence an entire protein using the Edman degradation or mass spectrometric methods. In practice, the Edman degradation is limited to peptides of 50 residues, because not all peptides in the reaction mixture release the amino acid derivative at each step. For instance, if the efficiency of release for each round were 98%, the proportion of "correct" amino acid released after 60 rounds would be (0.98^{60}), or 0.3—a hopelessly impure mix. Similarly, sequencing of long peptides by mass spectrometry yields a mass spectrum that can be complex and difficult to interpret. This obstacle can be circumvented by cleaving a protein into smaller peptides that can be sequenced. Protein cleavage can be achieved by chemical reagents, such as cyanogen bromide, or proteolytic enzymes, such as trypsin. Table 3.3 gives several other ways of specifically cleaving polypeptide chains. Note that these methods are sequence specific: they disrupt the protein backbone at particular amino acid residues in a predictable manner.

The peptides obtained by specific chemical or enzymatic cleavage are separated by some type of chromatography. The sequence of each purified

Table 3.3 Specific cleavage of polypeptides

Reagent	Cleavage site
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
O-Iodosobenzoate	Carboxyl side of tryptophan residues
Hydroxylamine	Asparagine–glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Clostripain	Carboxyl side of arginine residues
Staphylococcal protease	Carboxyl side of aspartate and glutamate residues (glutamate only under certain conditions)
Thrombin	Carboxyl side of arginine
Chymotrypsin	Carboxyl side of tyrosine, tryptophan, phenylalanine, leucine, and methionine
Carboxypeptidase A	Amino side of C-terminal amino acid (not arginine, lysine, or proline)

peptide is then determined by the methods described above. At this point, the amino acid sequences of segments of the protein are known, but the order of these segments is not yet defined. How can we order the peptides to obtain the primary structure of the original protein? The necessary additional information is obtained from *overlap peptides* (Figure 3.31). A second enzyme is used to split the polypeptide chain at different linkages. For example, chymotrypsin cleaves preferentially on the carboxyl side of aromatic and some other bulky nonpolar residues (Chapter 9). Because these chymotryptic peptides overlap two or more tryptic peptides, they can be used to establish the order of the peptides. The entire amino acid sequence of the polypeptide chain is then known.

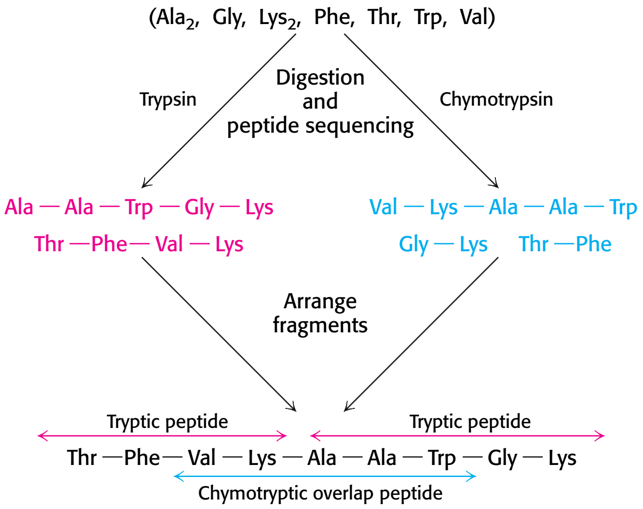

(Ala2, Gly, Lys2, Phe, Thr, Trp, Val)
Trypsin
Digestion and peptide sequencing
Chymotrypsin
Ala — Ala — Trp — Gly — Lys
Thr — Phe — Val — Lys
Val — Lys — Ala — Ala — Trp
Gly — Lys
Thr — Phe
Arrange fragments
Tryptic peptide
Tryptic peptide
Thr — Phe — Val — Lys — Ala — Ala — Trp — Gly — Lys
Chymotryptic overlap peptide

FIGURE 3.31 Overlap peptides. The peptide obtained by chymotryptic digestion overlaps two tryptic peptides, establishing their order.

Additional steps are necessary if the initial protein sample is actually several polypeptide chains. SDS–gel electrophoresis under reducing conditions should display the number of chains. Alternatively, the number of distinct N-terminal amino acids could be determined. After a protein has been identified as being made up of two or more polypeptide chains, denaturing agents, such as urea or guanidine hydrochloride, are used to dissociate chains held together by noncovalent bonds. The dissociated chains must be separated from one another before sequence determination can begin. Polypeptide chains linked by disulfide bonds are separated by reduction with thiols such as β-mercaptoethanol or dithiothreitol. To prevent the cysteine residues from recombining, they are alkylated with iodoacetate to form stable *S*-carboxymethyl derivatives (Figure 3.32). Sequencing can then be performed as already described.

Genomic and proteomic methods are complementary

Despite the technological advancements in both chemical and mass spectrometric methods of peptide sequencing, heroic effort is required to elucidate the sequence of large proteins, those with more than 1000 residues. For sequencing such proteins, a complementary experimental approach based on recombinant DNA technology is often more efficient. As will be discussed in Chapter 5, long stretches of DNA can be cloned and sequenced, and the nucleotide sequence can be translated to reveal the amino acid sequence of the protein encoded by the gene (Figure 3.33). Recombinant

DNA technology is producing a wealth of amino acid sequence information at a remarkable rate.

Nevertheless, even with the use of the DNA base sequence to determine primary structure, there is still a need to work with isolated proteins. The amino acid sequence deduced by reading the DNA sequence is that of the nascent protein, the direct product of the translational machinery. However, many proteins undergo *posttranslational modifications* after their syntheses. Some have their ends trimmed, and others arise by cleavage of a larger initial polypeptide chain. Cysteine residues in some proteins are oxidized to form disulfide links, connecting either parts within a chain or separate polypeptide chains. Specific side chains of some proteins are altered. Amino acid sequences derived from DNA sequences are rich in information, but they do not disclose these modifications. Chemical analyses of proteins in their mature form are needed to delineate the nature of these changes, which are critical for the biological activities of most proteins. *Thus, genomic and proteomic analyses are complementary approaches to elucidating the structural basis of protein function.*

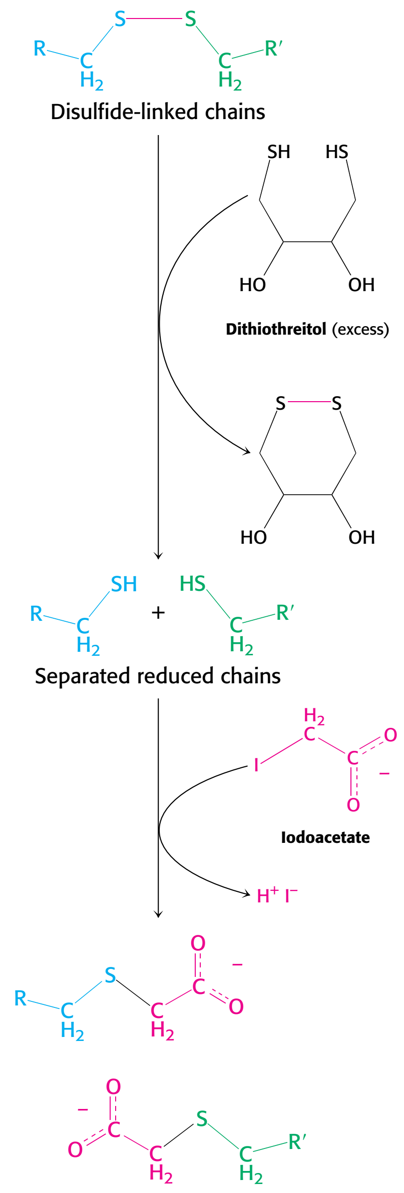


FIGURE 3.32 Disulfide-bond reduction. Polypeptides linked by disulfide bonds can be separated by reduction with dithiothreitol followed by alkylation to prevent them from re-forming.

The amino acid sequence of a protein provides valuable information

Regardless of the method used for its determination, the amino acid sequence of a protein can provide the biochemist with a wealth of information as to the protein's structure, function, and history.

1. *The sequence of a protein of interest can be compared with all other known sequences to ascertain whether significant similarities exist.* A search for kinship between a newly sequenced protein and the millions of previously sequenced ones takes only a few seconds on a personal computer (Chapter 6). If the newly isolated protein is a member of an established class of protein, we can begin to infer information about the protein's structure and function. For instance, chymotrypsin and trypsin are members of the serine protease family, a clan of proteolytic enzymes that have a common catalytic mechanism based on a reactive serine residue (Chapter 9). If the sequence of the newly isolated protein shows sequence similarity with trypsin or chymotrypsin, the result suggests that it may be a serine protease.

2. *Comparison of sequences of the same protein in different species yields a wealth of information about evolutionary pathways.* Genealogical relationships between species can be inferred from sequence differences between their proteins. If we assume that the random mutation rate of proteins over time is constant, then careful sequence comparison of related proteins between two organisms can provide an estimate for when these two evolutionary lines diverged. For example, a comparison of serum albumins found in primates indicates that human beings and African apes diverged 5 million years ago, not 30 million years ago as was once thought. Sequence analyses have opened a new perspective on the fossil record and the pathway of human evolution.

DNA sequence	GGG	TTC	TTG	GGA	GCA	GCA	GGA	AGC	ACT	ATG	GGC	GCA
Amino acid sequence	Gly	Phe	Leu	Gly	Ala	Ala	Gly	Ser	Thr	Met	Gly	Ala

FIGURE 3.33 DNA sequence yields the amino acid sequence. The complete nucleotide sequence of HIV-1 (human immunodeficiency virus), the cause of AIDS (acquired immune deficiency syndrome), was determined within a year after the isolation of the virus. A part of the DNA sequence specified by the RNA genome of the virus is shown here with the corresponding amino acid sequence (deduced from a knowledge of the genetic code).

3. *Amino acid sequences can be searched for the presence of internal repeats*. Such internal repeats can reveal the history of an individual protein itself. Many proteins apparently have arisen by duplication of primordial genes followed by their diversification. For example, calmodulin, a ubiquitous calcium sensor in eukaryotes, contains four similar calcium-binding modules that arose by gene duplication (Figure 3.34).

4. *Many proteins contain amino acid sequences that serve as signals designating their destinations or controlling their processing*. For example, a protein destined for export from a cell or for location in a membrane contains a *signal sequence,* a stretch of about 20 hydrophobic residues near the amino terminus that directs the protein to the appropriate membrane. Another protein may contain a stretch of amino acids that functions as a *nuclear localization signal,* directing the protein to the nucleus.

5. *Sequence data provide a basis for preparing antibodies specific for a protein of interest*. One or more parts of the amino acid sequence of a protein will elicit an antibody when injected into a mouse or rabbit. These specific antibodies can be very useful in determining the amount of a protein present in solution or in the blood, ascertaining its distribution within a cell, or cloning its gene (Section 3.2).

6. *Amino acid sequences are valuable for making DNA probes that are specific for the genes encoding the corresponding proteins*. Knowledge of a protein's primary structure permits the use of reverse genetics. DNA sequences that correspond to a part of the amino acid sequence can be constructed on the basis of the genetic code. These DNA sequences can be used as probes to isolate the gene encoding the protein so that the entire sequence of the protein can be determined. The gene in turn can provide valuable information about the physiological regulation of the protein. Protein sequencing is an integral part of molecular genetics, just as DNA cloning is central to the analysis of protein structure and function. We will revisit some of these topics in more detail in Chapter 5.

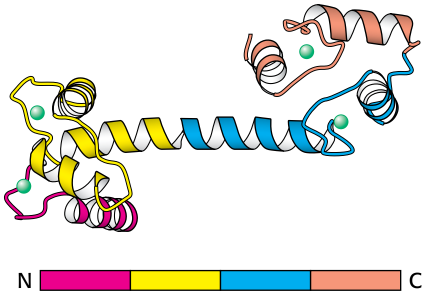


FIGURE 3.34 Repeating motifs in a protein chain. Calmodulin, a calcium sensor, contains four similar units (shown in red, yellow, blue, and orange) in a single polypeptide chain. *Notice* that each unit binds a calcium ion (shown in green). [Drawn from 1CLL.pdb.]

Individual proteins can be identified by mass spectrometry

The combination of mass spectrometry with chromatographic and peptide-cleavage techniques enables highly sensitive protein identification in complex biological mixtures. When a protein is cleaved by chemical or enzymatic methods (Table 3.3), a specific and predictable family of peptide fragments is formed. We learned in Chapter 2 that each protein has a unique, precisely defined amino acid sequence. Hence, the identity of the individual peptides formed from this cleavage reaction—and, importantly, their corresponding masses—is a distinctive signature for that particular protein. Protein cleavage, followed by chromatographic separation and mass spectrometry, enables rapid identification and quantitation of these signatures, even if they are present at very low concentrations. This technique for protein identification is referred to as *peptide mass fingerprinting*.

The speed and sensitivity of mass spectrometry has made this technology critical for the study of proteomics. Let us consider the analysis of the nuclear-pore complex from yeast, which facilitates the transport of large molecules into and out of the nucleus. This huge macromolecular complex was purified from yeast cells by careful procedures. The purified complex was fractionated by HPLC followed by gel electrophoresis. Individual bands from the gel were isolated, cleaved with trypsin, and analyzed by MALDI-TOF mass spectrometry. The fragments produced were compared with amino acid sequences deduced from the DNA sequence of the yeast genome as shown in Figure 3.35. A total of 174 nuclear-pore proteins were identified

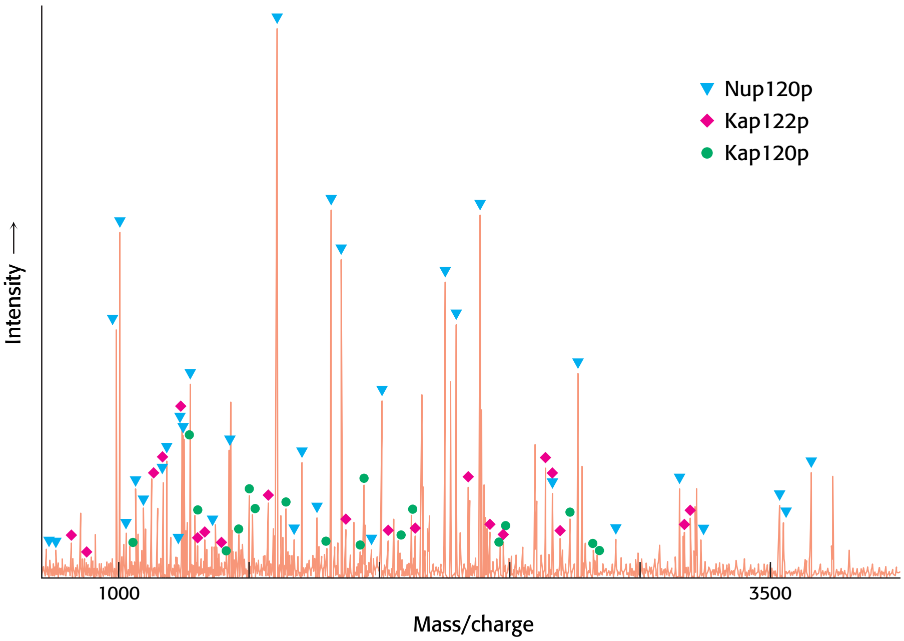


FIGURE 3.35 Proteomic analysis by mass spectrometry. This mass spectrum was obtained by analyzing a trypsin-treated band in a gel derived from a yeast nuclear-pore sample. Many of the peaks were found to match the masses predicted for peptide fragments from three proteins (Nup120p, Kap122p, and Kap120p) within the yeast genome. The band corresponded to an apparent molecular mass of 100 kDa. [Data from M. P. Rout, J. D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, and B. T. Chait. *J. Cell Biol.* 148:635–651, 2000.]

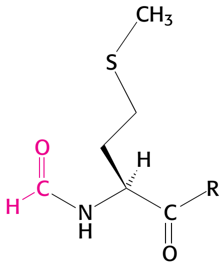
in this manner. Many of these proteins had not previously been identified as being associated with the nuclear pore despite years of study. Furthermore, mass spectrometric methods are sensitive enough to detect essentially all components of the pore if they are present in the samples used. Thus, a complete list of the components constituting this macromolecular complex could be obtained in a straightforward manner. Proteomic analysis of this type is growing in power as mass spectrometric and biochemical fractionation methods are refined.

3.4 Peptides Can Be Synthesized by Automated Solid-Phase Methods

Peptides of defined sequence can be synthesized to assist in biochemical analysis. These peptides are valuable tools for several purposes.

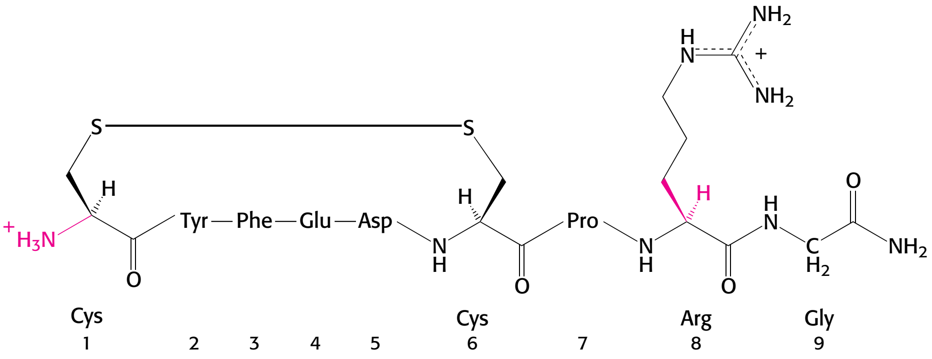
1. *Synthetic peptides can serve as antigens to stimulate the formation of specific antibodies*. Suppose we want to isolate the protein expressed by a specific gene. Peptides can be synthesized that match the translation of part of the gene's nucleic acid sequence, and antibodies can be generated that target these peptides. These antibodies can then be used to isolate the intact protein or localize it within the cell.

2. *Synthetic peptides can be used to isolate receptors for many hormones and other signal molecules*. For example, white blood cells are attracted to bacteria by formylmethionyl (fMet) peptides released in the breakdown of bacterial proteins. Synthetic formylmethionyl peptides have been useful in identifying the cell-surface receptor for this class of peptide. Moreover, synthetic peptides can be attached to agarose beads to prepare affinity chromatography columns for the purification of receptor proteins that specifically recognize the peptides.

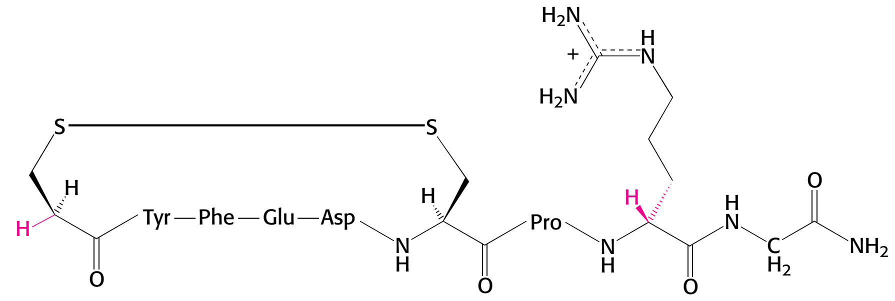


fMet peptide

3. *Synthetic peptides can serve as drugs.* Vasopressin is a peptide hormone that stimulates the reabsorption of water in the distal tubules of the kidney, leading to the formation of more-concentrated urine. Patients with diabetes insipidus are deficient in vasopressin (also called *antidiuretic hormone*), and so they excrete large volumes of dilute urine (more than 5 liters per day)



(A) **8-Arginine vasopressin**
(antidiuretic hormone, ADH)



(B) **1-Desamino-8-D-arginine vasopressin**

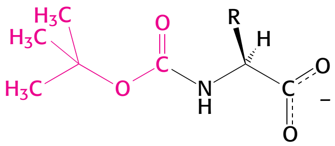
FIGURE 3.36 Vasopressin and a synthetic vasopressin analog. Structural formulas of (A) vasopressin, a peptide hormone that stimulates water resorption, and (B) 1-desamino-8-D-arginine vasopressin, a more stable synthetic analog of this antidiuretic hormone.

and are continually thirsty. This defect can be treated by administering 1-desamino-8-D-arginine vasopressin, a synthetic analog of the missing hormone (Figure 3.36). This synthetic peptide is degraded in vivo much more slowly than vasopressin and does not increase blood pressure.

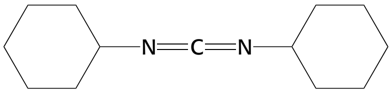
4. Finally, *studying synthetic peptides can help define the rules governing the three-dimensional structure of proteins*. We can ask whether a particular sequence by itself tends to fold into an α helix, a β strand, or a hairpin turn or behaves as a random coil. The peptides created for such studies can incorporate amino acids not normally found in proteins, allowing more variation in chemical structure than is possible with the use of only 20 amino acids.

How are these peptides constructed? The amino group of one amino acid is linked to the carboxyl group of another. However, a unique product is formed only if a single amino group and a single carboxyl group are available for reaction. Therefore, it is necessary to block some groups and to activate others to prevent unwanted reactions. First, the carboxyl-terminal amino acid is attached to an insoluble resin by its carboxyl group, effectively protecting it from further peptide-bond-forming reactions (Figure 3.37). The α-amino group of this amino acid is blocked with a protecting group such as a *tert*-butyloxycarbonyl (*t*-Boc) group. The *t*-Boc protecting group of this amino acid is then removed with trifluoroacetic acid.

The next amino acid (in the protected *t*-Boc form) and *dicyclohexylcarbodiimide* (DCC) are added together. At this stage, only the carboxyl group of the incoming amino acid and the amino group of the resin-bound amino acid are free to form a peptide bond. DCC reacts with the carboxyl group of the incoming amino acid, activating it for the peptide-bond-forming reaction. After the peptide bond has formed, excess reagents are washed away, leaving the desired dipeptide product attached to the beads. Additional amino acids are linked by the same sequence of reactions. At the end of the synthesis, the peptide is released from the beads by the



***t*-Butyloxycarbonyl amino acid**
(*t*-Boc amino acid)



Dicyclohexylcarbodiimide
(DCC)

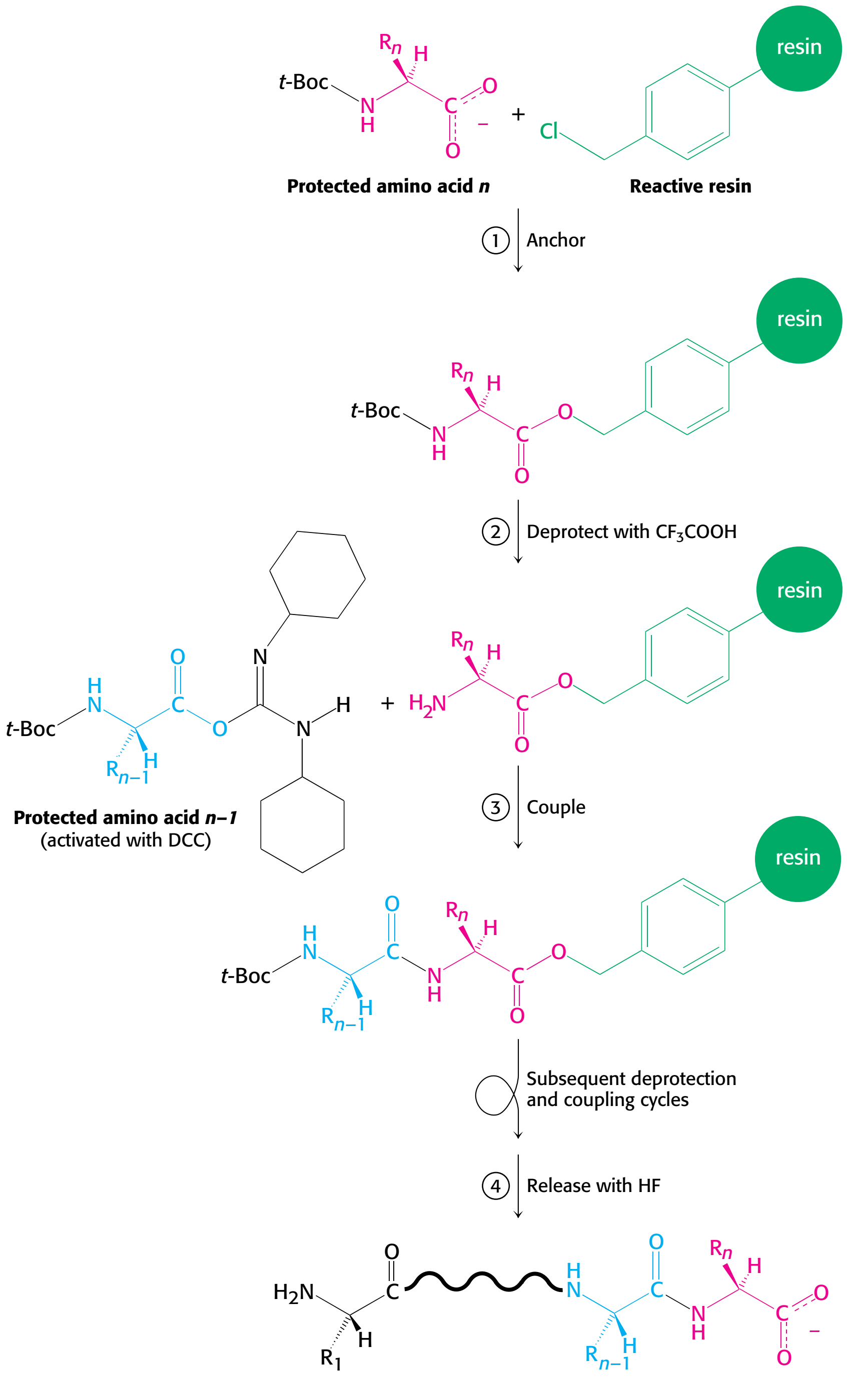

resin
t-Boc
Protected amino acid *n*
Reactive resin
① Anchor
② Deprotect with CF_3COOH
Protected amino acid *n*−*1*
(activated with DCC)
③ Couple
Subsequent deprotection and coupling cycles
④ Release with HF
H_2N
R_n
R_{n-1}
R_1

FIGURE 3.37 Solid-phase peptide synthesis. The sequence of steps in solid-phase synthesis is: (1) anchoring of the C-terminal amino acid to a solid resin, (2) deprotection of the amino terminus, and (3) coupling of the free amino terminus with the DCC-activated carboxyl group of the next amino acid. Steps 2 and 3 are repeated for each added amino acid. Finally, in step 4, the completed peptide is released from the resin.

addition of hydrofluoric acid (HF), which cleaves the carboxyl ester anchor without disrupting peptide bonds. Protecting groups on potentially reactive side chains, such as that of lysine, also are removed at this time.

A major advantage of this *solid-phase method*, first developed by R. Bruce Merrifield, is that the desired product at each stage is bound to beads that can be rapidly filtered and washed. Hence, there is no need to purify intermediates. All reactions are carried out in a single vessel, eliminating losses caused by repeated transfers of products. This cycle of reactions can be readily automated, which makes it feasible to routinely synthesize peptides containing about 50 residues in good yield and purity. In fact, the solid-phase method has been used to synthesize interferons (155 residues) that have antiviral activity and ribonuclease (124 residues) that is catalytically active. The protecting groups and cleavage agents may be varied for increased flexibility or convenience.

Synthetic peptides can be linked to create even longer molecules. With the use of specially developed *peptide-ligation* methods, proteins of 100 amino acids or more can by synthesized in very pure form. These methods enable the construction of even sharper tools for examining protein structure and function.

3.5 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy

Elucidation of the three-dimensional structure of a protein is often the source of a tremendous amount of insight into its corresponding function, inasmuch as the specificity of active sites and binding sites is defined by the precise atomic arrangement within these regions. For example, knowledge of the structure of a protein enables the biochemist to predict its mechanism of action, the effects of mutations on its function, and the desired features of drugs that may inhibit or augment its activity. X-ray crystallography and nuclear magnetic resonance spectroscopy are the two most important techniques for elucidating the conformation of proteins.

X-ray crystallography reveals three-dimensional structure in atomic detail

X-ray crystallography was the first method developed to determine protein structure in atomic detail. This technique provides the clearest visualization of the precise three-dimensional positions of most atoms within a protein. Of all forms of radiation, x-rays provide the best resolution for the determination of molecular structures because their wavelength approximately corresponds to the length of a covalent bond. The three components in an x-ray crystallographic analysis are a *protein crystal,* a *source of x-rays,* and a *detector* (Figure 3.38).

X-ray crystallography first requires the preparation of a protein or protein complex in crystal form, in which all protein molecules are oriented in a fixed, repeated arrangement with respect to one another. Slowly adding ammonium sulfate or another salt to a concentrated solution of protein to reduce its solubility favors the formation of highly ordered crystals—the process of salting out discussed on page 68. For example, myoglobin crystallizes in 3 M ammonium sulfate. Protein crystallization can be quite challenging: a concentrated solution of highly pure material is required and it is often difficult to predict which experimental conditions will yield the most-effective crystals. Methods for screening many different crystallization conditions using a small amount of protein sample have been developed. Typically, hundreds of conditions must be tested to obtain crystals fully suitable for crystallographic studies. Nevertheless, increasingly large and complex proteins have been crystallized. For example, poliovirus, an 8500-kDa assembly of 240 protein subunits surrounding an RNA core, has been crystallized and its structure solved by x-ray methods. Crucially, proteins frequently crystallize in their biologically active configuration. Enzyme crystals may display catalytic activity if the crystals are suffused with substrate.

After a suitably pure crystal of protein has been obtained, a source of x-rays is required. A beam of x-rays of wavelength 1.54 Å is produced by accelerating electrons against a copper target. Equipment suitable for generating x-rays in this manner is available in many laboratories. Alternatively, x-rays can be produced by *synchrotron radiation,* the acceleration of electrons in circular orbits at speeds close to the speed of light. Synchrotron-generated x-ray beams are much more intense than those generated by electrons hitting copper. The higher intensity enables the acquisition of high quality data from smaller crystals over a shorter exposure times. Several facilities throughout

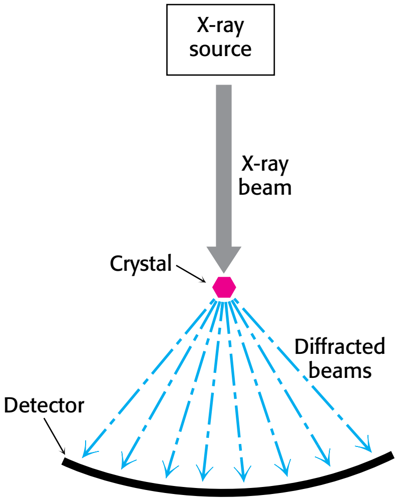


FIGURE 3.38 An x-ray crystallographic experiment. An x-ray source generates a beam, which is diffracted by a crystal. The resulting diffraction pattern is collected on a detector.

the world generate synchrotron radiation, such as the Advanced Light Source at Argonne National Laboratory outside Chicago and the Photon Factory in Tsukuba City, Japan.

When a narrow beam of x-rays is directed at the protein crystal, most of the beam passes directly through the crystal while a small part is scattered in various directions. These scattered, or *diffracted,* x-rays can be detected by x-ray film or by a solid-state electronic detector. The scattering pattern provides abundant information about protein structure. The basic physical principles underlying the technique are:

1. *Electrons scatter x-rays.* The amplitude of the wave scattered by an atom is proportional to its number of electrons. Thus, a carbon atom scatters six times as strongly as a hydrogen atom does.
2. *The scattered waves recombine.* Each diffracted beam comprises waves scattered by each atom in the crystal. The scattered waves reinforce one another at the film or detector if they are in phase (in step) there, and they cancel one another if they are out of phase.
3. *The way in which the scattered waves recombine depends only on the atomic arrangement.*

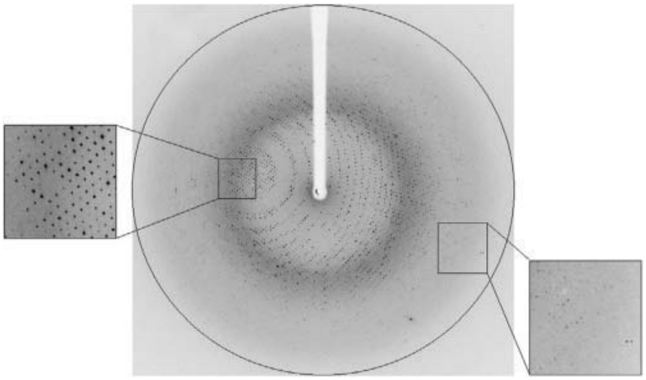


FIGURE 3.39 An x-ray diffraction pattern. A protein crystal diffracts x-rays to produce a pattern of spots, or reflections, on the detector surface. The white silhouette in the center of the image is from a beam stop which protects the detector from the intense, undiffracted x-rays. [S. Lansky, O. Alalouf, V. Solomon, A. Alhassid, L. Govada, N. E. Chayan, H. Belrhali, Y. Shoham, and G. Shoham. *Acta Cryst*. F69:430–434, 2013, Fig. 2.]

The protein crystal is mounted and positioned in a precise orientation with respect to the x-ray beam and the film. The crystal is rotated so that the beam can strike the crystal from many directions. This rotational motion results in an x-ray photograph consisting of a regular array of spots called *reflections*. The x-ray photograph shown in Figure 3.39 is a two-dimensional section through a three-dimensional array of 72,000 reflections. The intensities and positions of these reflections are the basic experimental data of an x-ray crystallographic analysis. Each reflection is formed from a wave with an amplitude proportional to the square root of the observed intensity of the spot. Each wave also has a *phase*—that is, the timing of its crests and troughs relative to those of other waves. Additional experiments or calculations must be performed to determine the phases corresponding to each reflection.

The next step is to reconstruct an image of the protein from the observed reflections. In light microscopy or electron microscopy, the diffracted beams are focused by lenses to directly form an image. However, appropriate lenses for focusing x-rays do not exist. Instead, the image is formed by applying a mathematical relation called a *Fourier transform* to the measured amplitudes and calculated phases of every observed reflection. The image obtained is referred to as the *electron-density map*. It is a three-dimensional graphic representation of where the electrons are most densely localized and is used to determine the positions of the atoms in the crystallized molecule (Figure 3.40). Critical to the interpretation of the map is its *resolution,* which is determined by the number of scattered intensities used in the Fourier transform. The fidelity of the image depends on this resolution, as shown by the optical analogy in Figure 3.41.

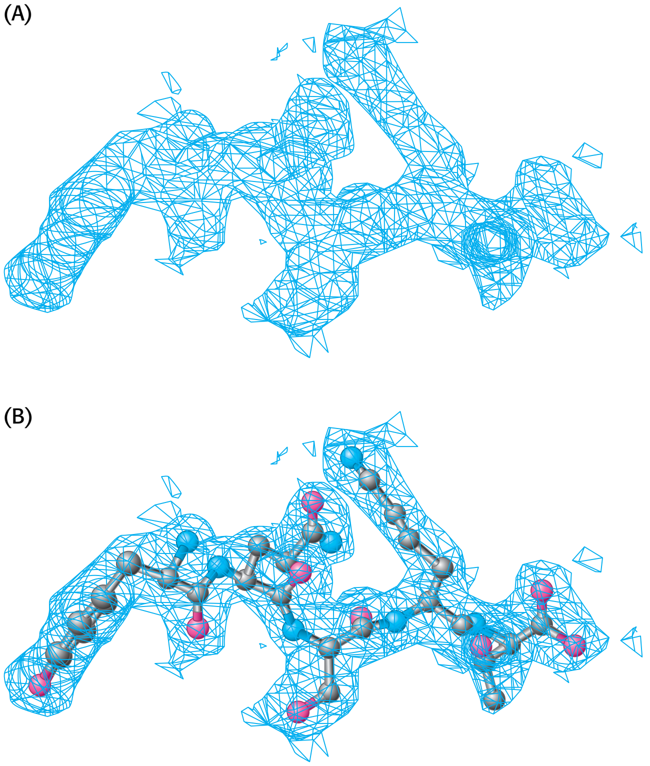


Figure 3.40 Interpretation of an electron-density map. (A) A segment of an electron-density map is drawn as a three-dimensional contour plot, in which the regions inside the "cage" represent the regions of highest electron density. (B) A model of the protein is built into this map so as to maximize the placement of atoms within this density. [Drawn from 1FCH.pdb.]

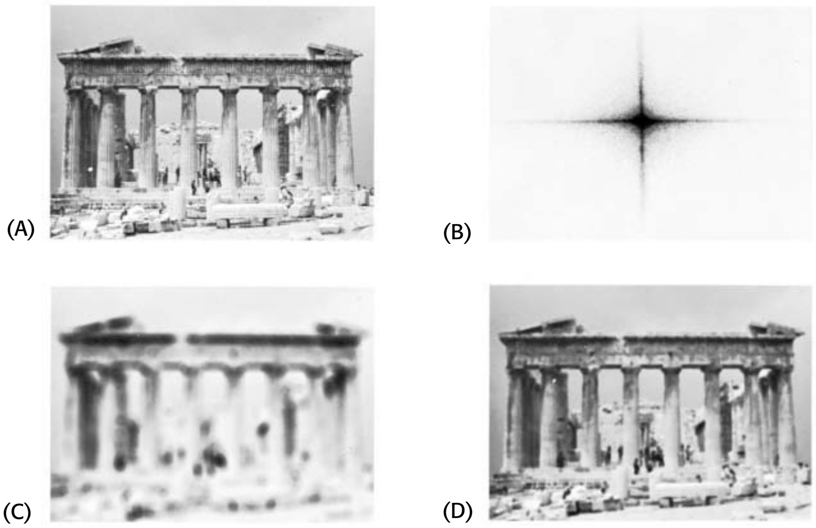


FIGURE 3.41 Resolution affects the quality of an image. The effect of resolution on the quality of a reconstructed image is shown by an optical analog of x-ray diffraction: (A) a photograph of the Parthenon; (B) an optical diffraction pattern of the Parthenon; (C and D) images reconstructed from the pattern in part B. More data were used to obtain image D than image C, which accounts for the higher quality of image D. [Courtesy of Dr. Thomas Steitz (part A) and Dr. David DeRosier (part B).]

A resolution of 6 Å reveals the course of the polypeptide chain but few other structural details. The reason is that polypeptide chains pack together so that their centers are between 5 Å and 10 Å apart. Maps at higher resolution are needed to delineate groups of atoms, which lie between 2.8 Å and 4.0 Å apart, and individual atoms, which are between 1.0 Å and 1.5 Å apart (Figure 3.42). The ultimate resolution of an x-ray analysis is determined by the degree of perfection of the crystal. For proteins, this limiting resolution is often about 2 Å; however, in exceptional cases, resolutions of 1.0 Å have been obtained.

Nuclear magnetic resonance spectroscopy can reveal the structures of proteins in solution

X-ray crystallography is the most powerful method for determining protein structures. However, some proteins do not readily crystallize. Furthermore, although structures of crystallized proteins very closely represent those of proteins free of the constraints imposed by the crystalline environment, structures in solution can be sources of additional insights. *Nuclear magnetic resonance* (NMR) *spectroscopy* is unique in being able to reveal the atomic structure of macromolecules *in solution,* provided that highly concentrated solutions

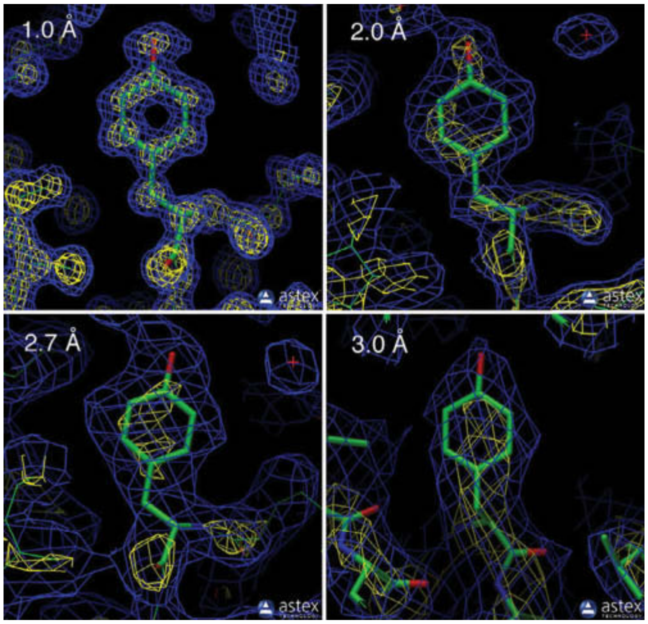


FIGURE 3.42 Resolution affects the quality of the electron density map. The electron density maps of a tyrosine residue at four different resolution levels (1.0 Å, 2.0 Å, 2.7 Å, and 3.0 Å) are shown. At the lower resolution levels (2.7 Å and 3.0 Å), only a group of atoms corresponding to the side chain is visible, whereas at the highest resolution (1.0 Å), individual atoms within the side chain are distinguishable. [Data from www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/resolution.html]

TABLE 3.4 Biologically important nuclei giving NMR signals

Nucleus	Natural abundance (% by weight of the element)
^{1}H	99.984
^{2}H	0.016
^{13}C	1.108
^{14}N	99.635
^{15}N	0.365
^{17}O	0.037
^{23}Na	100.0
^{25}Mg	10.05
^{31}P	100.0
^{35}Cl	75.4
^{39}K	93.1

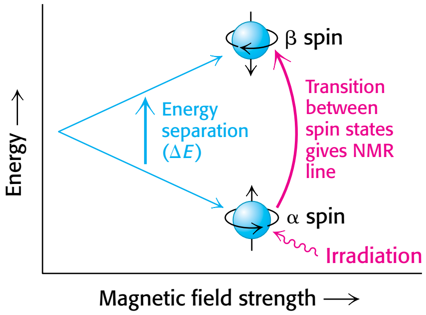

β spin
Transition between spin states gives NMR line
Energy separation (ΔE)
Energy →
α spin
Irradiation
Magnetic field strength →

FIGURE 3.43 Basis of NMR spectroscopy. The energies of the two orientations of a nucleus of spin $^{1}/_{2}$ (such as ^{31}P and ^{1}H) depend on the strength of the applied magnetic field. Absorption of electromagnetic radiation of appropriate frequency induces a transition from the lower to the upper level.

(~1 mM, or 15 mg ml^{-1} for a 15-kDa protein) can be obtained. This technique depends on the fact that certain atomic nuclei are intrinsically magnetic. Only a limited number of isotopes display this property, called *spin,* and those most important to biochemistry are listed in Table 3.4. The simplest example is the hydrogen nucleus (^{1}H), which is a proton. The spinning of a proton generates a magnetic moment. This moment can take either of two orientations, or spin states (called α and β), when an external magnetic field is applied (Figure 3.43). The energy difference between these states is proportional to the strength of the imposed magnetic field. The α state has a slightly lower energy because it is aligned with this applied field. Hence, in a given population of nuclei, slightly more will occupy the α state (by a factor of the order of 1.00001 in a typical experiment). A spinning proton in an α state can be raised to an excited state (β state) by applying a pulse of electromagnetic radiation (a radio-frequency, or RF, pulse), provided that the frequency corresponds to the energy difference between the α and the β states. In these circumstances, the spin will change from α to β; in other words, *resonance* will be obtained.

These properties can be used to examine the chemical surroundings of the hydrogen nucleus. The flow of electrons around a magnetic nucleus generates a small local magnetic field that opposes the applied field. The degree of such shielding depends on the surrounding electron density. Consequently, nuclei in different environments will change states, or resonate, at slightly different field strengths or radiation frequencies. A resonance spectrum for a molecule is obtained by keeping the magnetic field constant and varying the frequency of the electromagnetic radiation. The nuclei of the perturbed sample absorb electromagnetic radiation at a frequency that can be measured. The different frequencies, termed *chemical shifts,* are expressed in fractional units δ (parts per million, or ppm) relative to the shifts of a standard compound, such as a water-soluble derivative of tetramethylsilane, that is added with the sample. For example, a $—CH_3$ proton typically exhibits a chemical shift (δ) of 1 ppm, compared with a chemical shift of 7 ppm for an aromatic proton. The chemical shifts of most protons in protein molecules fall between 0 and 9 ppm (Figure 3.44). Most protons in many proteins can be resolved by using this technique of *one-dimensional NMR.* With this information, we can then deduce changes to a particular chemical group under different conditions, such as the conformational change of a protein from a disordered structure to an α helix in response to a change in pH.

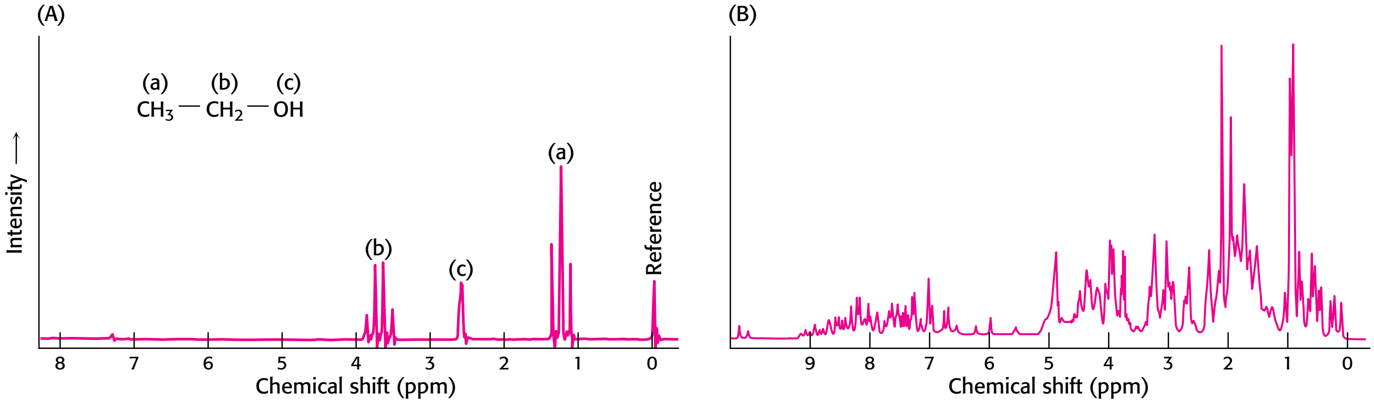


FIGURE 3.44 One-dimensional NMR spectra. (A) ^{1}H-NMR spectrum of ethanol (CH_3CH_2OH) shows that the chemical shifts for the hydrogen are clearly resolved. (B) ^{1}H-NMR spectrum of a 55 amino acid fragment of a protein having a role in RNA splicing shows a greater degree of complexity. A large number of peaks are present and many overlap. [(A) Data from C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland, 1991), p. 280; (B) courtesy of Dr. Barbara Amann and Dr. Wesley McDermott.]

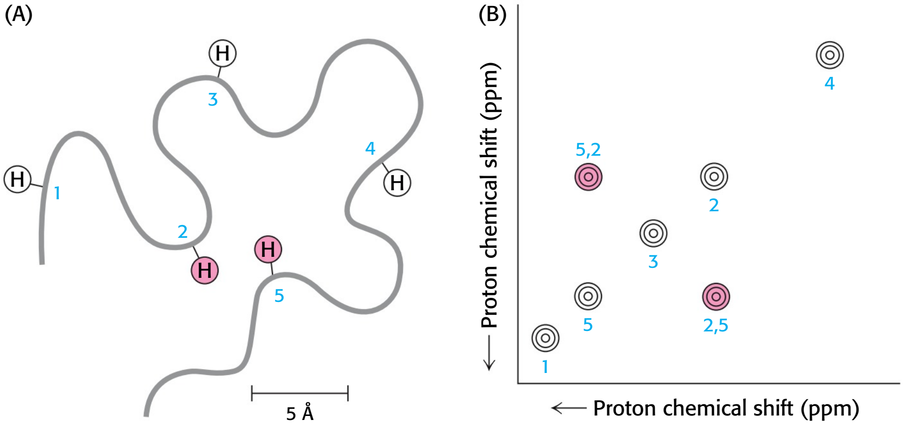


FIGURE 3.45 The nuclear Overhauser effect. The nuclear Overhauser effect (NOE) identifies pairs of protons that are in close proximity. (A) Schematic representation of a polypeptide chain highlighting five particular protons. Protons 2 and 5 are in close proximity (~4 Å apart), whereas other pairs are farther apart. (B) A highly simplified NOESY spectrum. The diagonal shows five peaks corresponding to the five protons in part A. The peak above the diagonal and the symmetrically related one below reveal that proton 2 is close to proton 5.

We can garner even more information by examining how the spins on different protons affect their neighbors. By inducing a transient magnetization in a sample through the application of a radio-frequency pulse, we can alter the spin on one nucleus and examine the effect on the spin of a neighboring nucleus. Especially revealing is a *two-dimensional spectrum obtained by nuclear Overhauser enhancement spectroscopy* (NOESY), *which graphically displays pairs of protons that are in close proximity,* even if they are not close together in the primary structure. The basis for this technique is the *nuclear Overhauser effect* (NOE), an interaction between nuclei that is proportional to the inverse sixth power of the distance between them. Magnetization is transferred from an excited nucleus to an unexcited one if the two nuclei are less than about 5 Å apart (Figure 3.45A). In other words, the effect provides a means of detecting the location of atoms relative to one another in the three-dimensional structure of the protein. The peaks that lie along the diagonal of a NOESY spectrum (shown in white in Figure 3.45B) correspond to those present in a one-dimensional NMR experiment. The peaks apart from the diagonal (shown in red in Figure 3.45B), referred to as *off-diagonal peaks* or *cross-peaks,* provide crucial new information: *they identify pairs of protons that are less than 5 Å apart*. A two-dimensional NOESY spectrum for a protein comprising 55 amino acids is shown in Figure 3.46. The large number of off-diagonal peaks reveals short proton–proton distances. The three-dimensional structure of a protein can be

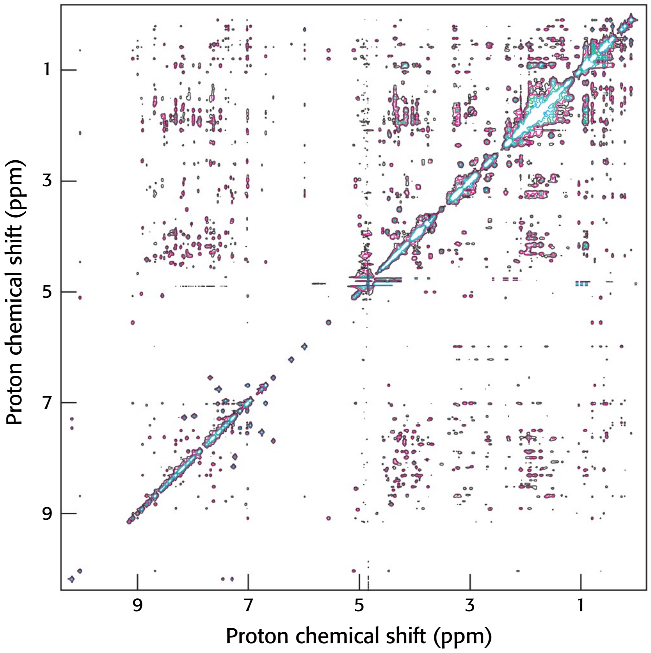


FIGURE 3.46 Detecting short proton–proton distances. A NOESY spectrum for a 55 amino acid domain from a protein having a role in RNA splicing. Each off-diagonal peak corresponds to a short proton–proton separation. This spectrum reveals hundreds of such short proton–proton distances, which can be used to determine the three-dimensional structure of this domain. [Courtesy of Dr. Barbara Amann and Dr. Wesley McDermott.]

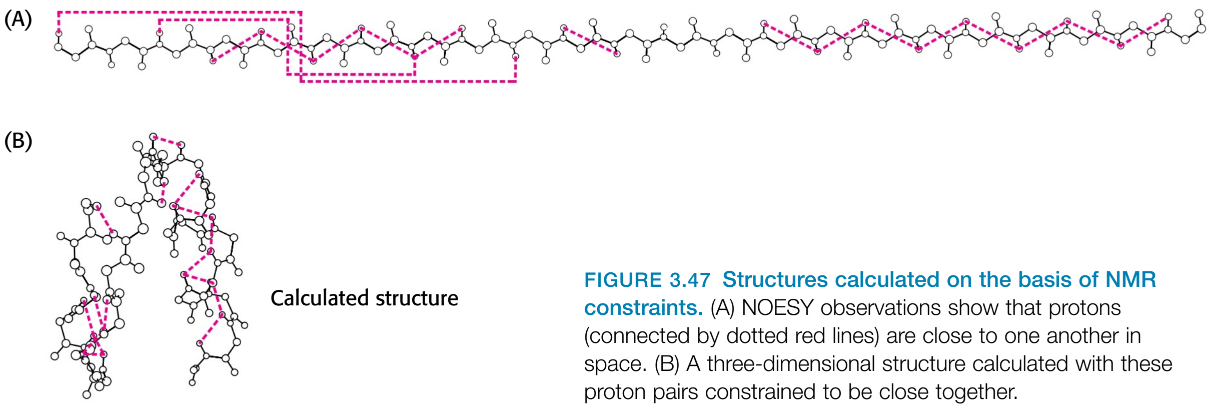


FIGURE 3.47 Structures calculated on the basis of NMR constraints. (A) NOESY observations show that protons (connected by dotted red lines) are close to one another in space. (B) A three-dimensional structure calculated with these proton pairs constrained to be close together.

reconstructed with the use of such proximity relations. Structures are calculated such that protons that must be separated by less than 5 Å on the basis of NOESY spectra are close to one another in the three-dimensional structure (Figure 3.47). If a sufficient number of distance constraints are applied, the three-dimensional structure can nearly be determined uniquely.

In practice, a family of related structures is generated by NMR spectroscopy for three reasons (Figure 3.48). First, not enough constraints may be experimentally accessible to fully specify the structure. Second, the distances obtained from analysis of the NOESY spectrum are only approximate. Finally, the experimental observations are made not on single molecules but on a large number of molecules in solution that may have slightly different structures at any given moment. Thus, the family of structures generated from NMR structure analysis indicates the range of conformations for the protein in solution. At present, structural determination by NMR spectroscopy is generally limited to proteins less than 50 kDa, but its resolving power is certain to increase. The power of NMR has been greatly enhanced by the ability of recombinant DNA technology to produce proteins labeled uniformly or at specific sites with ^{13}C, ^{15}N, and ^{2}H (Chapter 5).

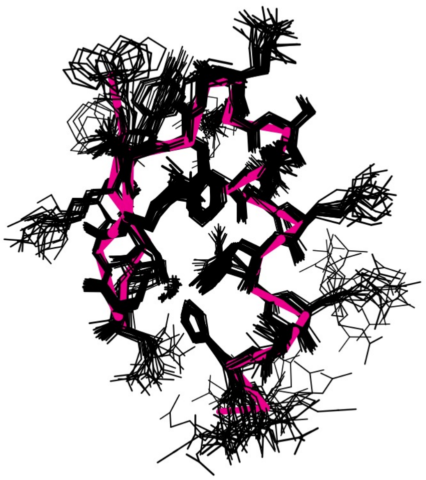


FIGURE 3.48 A family of structures. A set of 25 structures for a 28 amino acid domain from a zinc-finger-DNA-binding protein. The red line traces the average course of the protein backbone. Each of these structures is consistent with hundreds of constraints derived from NMR experiments. The differences between the individual structures are due to a combination of imperfections in the experimental data and the dynamic nature of proteins in solution. [Courtesy of Dr. Barbara Amann.]

The structures of nearly 97,000 proteins had been elucidated by x-ray crystallography and NMR spectroscopy by the end of 2013 and several new structures are now determined each day. The coordinates are collected at the Protein Data Bank (www.pdb.org), and the structures can be accessed for visualization and analysis. Knowledge of the detailed molecular architecture of proteins has been a source of insight into how proteins recognize and bind other molecules, how they function as enzymes, how they fold, and how they evolved. This extraordinarily rich harvest is continuing at a rapid pace and is greatly influencing the entire field of biochemistry as well as other biological and physical sciences.

SUMMARY

The rapid progress in gene sequencing has advanced another goal of biochemistry—elucidation of the proteome. The proteome is the complete set of proteins expressed and includes information about how they are modified, how they function, and how they interact with other molecules.

3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function

Proteins can be separated from one another and from other molecules on the basis of such characteristics as solubility, size, charge, and binding affinity. SDS–polyacrylamide gel electrophoresis separates the polypeptide chains of proteins under denaturing conditions largely according to mass. Proteins can also be separated electrophoretically on the basis of net charge by isoelectric focusing in a pH gradient. Ultracentrifugation and gel-filtration chromatography resolve proteins according to size, whereas ion-exchange chromatography separates them mainly on the basis of net charge. The high affinity of many proteins for specific chemical groups is exploited in affinity chromatography, in which proteins bind to columns containing beads bearing covalently linked substrates, inhibitors, or other specifically recognized groups. The mass of a protein can be determined by sedimentation-equilibrium measurements.

3.2 Immunology Provides Important Techniques with Which to Investigate Proteins

Proteins can be detected and quantitated by highly specific antibodies; monoclonal antibodies are especially useful because they are homogeneous. Enzyme-linked immunosorbent assays and western blots of SDS–polyacrylamide gels are used extensively. Proteins can also be localized within cells by immunofluorescence microscopy and immunoelectron microscopy.

3.3 Mass Spectrometry Is a Powerful Technique for the Identification of Peptides and Proteins

Techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) allow the generation of ions of proteins and peptides in the gas phase. The mass of such protein ions can be determined with great accuracy and precision. Masses determined by these techniques act as protein name tags because the mass of a protein or peptide is precisely determined by its amino acid composition and, hence, by its sequence. In addition to chemical methods, such as the Edman degradation, tandem mass spectrometry enables the rapid and highly accurate sequencing of peptides. These sequences are rich in information concerning the kinship of proteins, their evolutionary relationships, and diseases produced by mutations. Knowledge of a sequence provides valuable clues to conformation and function. Mass spectrometric techniques are central to proteomics because they make it possible to analyze the constituents of large macromolecular assemblies or other collections of proteins.

3.4 Peptides Can Be Synthesized by Automated Solid-Phase Methods

Polypeptide chains can be synthesized by automated solid-phase methods in which the carboxyl end of the growing chain is linked to an insoluble support. The carboxyl group of the incoming amino acid is activated by dicyclohexylcarbodiimide and joined to the amino group of the growing chain. Synthetic peptides can serve as drugs and as antigens to stimulate the formation of specific antibodies. They can also be sources of insight into the relation between amino acid sequence and conformation.

3.5 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy

X-ray crystallography and nuclear magnetic resonance spectroscopy have greatly enriched our understanding of how proteins fold, recognize other molecules, and catalyze chemical reactions. X-ray crystallography is possible because electrons scatter x-rays. The diffraction pattern produced can be analyzed to reveal the arrangement of atoms in a protein. The three-dimensional structures of tens of thousands of proteins are now known in atomic detail. Nuclear magnetic resonance spectroscopy reveals the structure and dynamics of proteins in solution. The chemical shift of nuclei depends on their local environment. Furthermore, the spins of neighboring nuclei interact with each other in ways that provide definitive structural information. This information can be used to determine complete three-dimensional structures of proteins.

KEY TERMS

proteome (p. 66)
assay (p. 67)
specific activity (p. 67)
homogenate (p. 67)
salting out (p. 68)
dialysis (p. 69)
gel-filtration chromatography (p. 69)
ion-exchange chromatography (p. 70)
cation exchange (p. 70)
anion exchange (p. 70)
affinity chromatography (p. 70)
high-performance liquid chromatography (HPLC) (p. 71)
gel electrophoresis (p. 71)
isoelectric point (p. 73)
isoelectric focusing (p. 73)
two-dimensional electrophoresis (p. 74)
sedimentation coefficient (Svedberg unit, S) (p. 76)
antibody (p. 79)
antigen (p. 79)
antigenic determinant (epitope) (p. 79)
polyclonal antibody (p. 80)
monoclonal antibody (p. 80)
enzyme-linked immunosorbent assay (ELISA) (p. 82)
western blotting (p. 83)
fluorescence microscopy (p. 84)
green fluorescent protein (GFP) (p. 84)
matrix-assisted laser desorption/ionization (MALDI) (p. 85)
electrospray ionization (ESI) (p. 85)
time-of-flight (TOF) mass analyzer (p. 85)
Edman degradation (p. 87)
phenyl isothiocyanate (p. 87)
tandem mass spectrometry (p. 87)
overlap peptide (p. 89)
peptide mass fingerprinting (p. 91)
solid-phase method (p. 94)
x-ray crystallography (p. 95)
Fourier transform (p. 96)
electron-density map (p. 96)
nuclear magnetic resonance (NMR) spectroscopy (p. 97)
chemical shift (p. 98)

PROBLEMS

1. *Valuable reagents.* The following reagents are often used in protein chemistry:

CNBr
Urea
Mercaptoethanol
Chymotrypsin
Trypsin
Performic acid
6 N HCl
Phenyl isothiocyanate

Which one is the best suited for accomplishing each of the following tasks?

(a) Determination of the amino acid sequence of a small peptide.

(b) Reversible denaturation of a protein devoid of disulfide bonds. Which additional reagent would you need if disulfide bonds were present?

(c) Hydrolysis of peptide bonds on the carboxyl side of aromatic residues.

(d) Cleavage of peptide bonds on the carboxyl side of methionines.

(e) Hydrolysis of peptide bonds on the carboxyl side of lysine and arginine residues.

2. *The only constant is change.* Explain how two different cell types from the same organism will have identical genomes but may have vastly divergent proteomes.

3. *Crafting a new breakpoint.* Ethyleneimine reacts with cysteine side chains in proteins to form *S*-aminoethyl derivatives. The peptide bonds on the carboxyl side of these modified cysteine residues are susceptible to hydrolysis by trypsin. Why?

4. *Spectrometry.* The absorbance *A* of a solution is defined as

$$A = \log_{10}(I_0/I)$$

in which I_0 is the incident-light intensity and I is the transmitted-light intensity. The absorbance is related to the molar absorption coefficient (extinction coefficient) ε (in M^{-1} cm^{-1}), concentration c (in M), and path length l (in cm) by

$$A = \varepsilon lc$$

The absorption coefficient of myoglobin at 580 nm is 15,000 M^{-1} cm^{-1}. What is the absorbance of a 1 mg ml^{-1} solution across a 1-cm path? What percentage of the incident light is transmitted by this solution?

5. *It's in the bag.* Suppose that you precipitate a protein with 1 M $(NH_4)_2SO_4$ and that you wish to reduce the concentration of the $(NH_4)_2SO_4$. You take 1 ml of your sample and dialyze it in 1000 ml of buffer. At the end of dialysis, what is the concentration of $(NH_4)_2SO_4$ in your sample? How could you further lower the $(NH_4)_2SO_4$ concentration?

6. *Too much or not enough.* Why do proteins precipitate at high salt concentrations? Although many proteins precipitate at high salt concentrations, some proteins require salt to dissolve in water. Explain why some proteins require salt to dissolve.

7. *A slow mover.* Tropomyosin, a 70-kDa muscle protein, sediments more slowly than does hemoglobin (65 kDa). Their sedimentation coefficients are 2.6S and 4.31S, respectively. Which structural feature of tropomyosin accounts for its slow sedimentation?

8. *Sedimenting spheres.* What is the dependence of the sedimentation coefficient *s* of a spherical protein on its mass? How much more rapidly does an 80-kDa protein sediment than does a 40-kDa protein?

9. *Frequently used in shampoos.* The detergent sodium dodecyl sulfate (SDS) denatures proteins. Suggest how SDS destroys protein structure.

10. *Size estimate.* The relative electrophoretic mobilities of a 30-kDa protein and a 92-kDa protein used as standards

on an SDS–polyacrylamide gel are 0.80 and 0.41, respectively. What is the apparent mass of a protein having a mobility of 0.62 on this gel?

11. *Unexpected migration.* Some proteins migrate anomalously in SDS-PAGE gels. For instance, the molecular weight determined from an SDS-PAGE gel is sometimes very different from the molecular weight determined from the amino acid sequence. Suggest an explanation for this discrepancy.

12. *Sorting cells.* Fluorescence-activated cell sorting (FACS) is a powerful technique for separating cells according to their content of particular molecules. For example, a fluorescence-labeled antibody specific for a cell-surface protein can be used to detect cells containing such a molecule. Suppose that you want to isolate cells that possess a receptor enabling them to detect bacterial degradation products. However, you do not yet have an antibody directed against this receptor. Which fluorescence-labeled molecule would you prepare to identify such cells?

13. *Column choice.* (a) The octapeptide AVGWRVKS was digested with the enzyme trypsin. Which method would be most appropriate for separating the products: ion-exchange or gel-filtration chromatography? Explain. (b) Suppose that the peptide was digested with chymotrypsin. What would be the optimal separation technique? Explain.

14. *Power(ful) tools.* Monoclonal antibodies can be conjugated to an insoluble support by chemical methods. Explain how these antibody-bound beads can be exploited for protein purification.

15. *Assay development.* You wish to isolate an enzyme from its native source and need a method for measuring its activity throughout the purification. However, neither the substrate nor the product of the enzyme-catalyzed reaction can be detected by spectroscopy. You discover that the product of the reaction is highly antigenic when injected into mice. Propose a strategy to develop a suitable assay for this enzyme.

16. *Making more enzyme?* In the course of purifying an enzyme, a researcher performs a purification step that results in an *increase* in the total activity to a value greater than that present in the original crude extract. Explain how the amount of total activity might increase.

17. *Divide and conquer.* The determination of the mass of a protein by mass spectrometry often does not allow its unique identification among possible proteins within a complete proteome, but determination of the masses of all fragments produced by digestion with trypsin almost always allows unique identification. Explain.

18. *Know your limits.* Which two amino acids are indistinguishable in peptide sequencing by the tandem mass spectrometry method described in this chapter and why?

19. *Protein purification problem.* Complete the following table.

Purification Procedure	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Purification level	Yield (%)
Crude extract	20,000	4,000,000		1	100
$(NH_4)_2SO_4$ precipitation	5,000	3,000,000			
DEAE-cellulose chromatography	1,500	1,000,000			
Gel-filtration chromatography	500	750,000			
Affinity chromatography	45	675,000			

20. *Part of the mix.* Your frustrated colleague hands you a mixture of four proteins with the following properties:

	Isoelectric point (pI)	Molecular weight (in kDa)
Protein A	4.1	80
Protein B	9.0	81
Protein C	8.8	37
Protein D	3.9	172

(a) Propose a method for the isolation of Protein B from the other proteins. (b) If Protein B also carried a His tag at its N-terminus, how could you revise your method?

21. *The challenge of flexibility.* Structures of proteins comprising domains separated by flexible linker regions can be quite difficult to solve by x-ray crystallographic methods. Why might this be the case? What are possible experimental approaches to circumvent this barrier?

Chapter Integration Problems

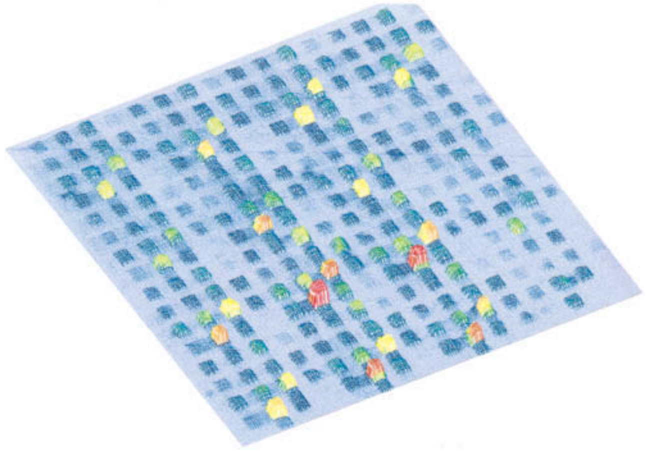
22. *Quaternary structure.* A protein was purified to homogeneity. Determination of the mass by gel-filtration chromatography yields 60 kDa. Chromatography in the presence of 6 M urea yields a 30-kDa species. When the chromatography is repeated in the presence of 6 M urea and 10 mM β-mercaptoethanol, a single molecular species of 15 kDa results. Describe the structure of the molecule.

23. *Helix–coil transitions.* (a) NMR measurements have shown that poly-L-lysine is a random coil at pH 7 but becomes a helix as the pH is raised above 10. Account for this pH-dependent conformational transition. (b) Predict the pH dependence of the helix–coil transition of poly-L-glutamate.

24. *Peptide mass determination.* You have isolated a protein from the bacterium *E. coli* and seek to confirm its identity by trypsin digestion and mass spectrometry. Determination of the masses of several peptide fragments has enabled you to deduce the identity of the protein. However, there is a discrepancy with one of the peptide fragments, which you believe should have the sequence MLNSFK and an $(M+H)^+$ value of 739.38. In your experiments, you repeatedly obtain an $(M+H)^+$ value of 767.38. What is the cause

of this discrepancy and what does it tell you about the region of the protein from which this peptide is derived?

25. *Peptides on a chip.* Large numbers of different peptides can be synthesized in a small area on a solid support. This high-density array can then be probed with a fluorescence-labeled protein to find out which peptides are recognized. The binding of an antibody to an array of 1024 different peptides occupying a total area the size of a thumbnail is shown in the adjoining illustration. How would you synthesize such a peptide array? (Hint: Use light instead of acid to deprotect the terminal amino group in each round of synthesis.)



Fluorescence scan of an array of 1024 peptides in a 1.6-cm^2 area. Each synthesis site is a 400-μm square. A fluorescently labeled monoclonal antibody was added to the array to identify peptides that are recognized. The height and color of each square denote the fluorescence intensity. [Information from S. P. A. Fodor et al., *Science* 251(1991):767.]

26. *Exchange rate.* The amide hydrogen atoms of peptide bonds within proteins can exchange with protons in the solvent. In general, amide hydrogen atoms in buried regions of proteins and protein complexes exchange more slowly than those on the solvent-accessible surface do. Determination of these rates can be used to explore the protein-folding reaction, probe the tertiary structure of proteins, and identify the regions of protein–protein interfaces. These exchange reactions can be followed by studying the behavior of the protein in solvent that has been labeled with deuterium (^{2}H), a stable isotope of hydrogen. What two methods described in this chapter could be readily applied to the study of hydrogen–deuterium exchange rates in proteins?

Data Interpretation Problems

27. *Protein sequencing 1.* Determine the sequence of hexapeptide on the basis of the following data. Note: When the sequence is not known, a comma separates the amino acids (Table 3.3).

Amino acid composition: (2R,A,S,V,Y)

N-terminal analysis of the hexapeptide: A

Trypsin digestion: (R,A,V) and (R,S,Y)

Carboxypeptidase digestion: No digestion.

Chymotrypsin digestion: (A,R,V,Y) and (R,S)

28. *Protein sequencing 2.* Determine the sequence of a peptide consisting of 14 amino acids on the basis of the following data.

Amino acid composition: (4S,2L,F,G,I,K,M,T,W,Y)

N-terminal analysis: S

Carboxypeptidase digestion: L

Trypsin digestion: (3S,2L,F,I,M,T,W) (G,K,S,Y)

Chymotrypsin digestion: (F,I,S) (G,K,L) (L,S) (M,T) (S,W) (S,Y)

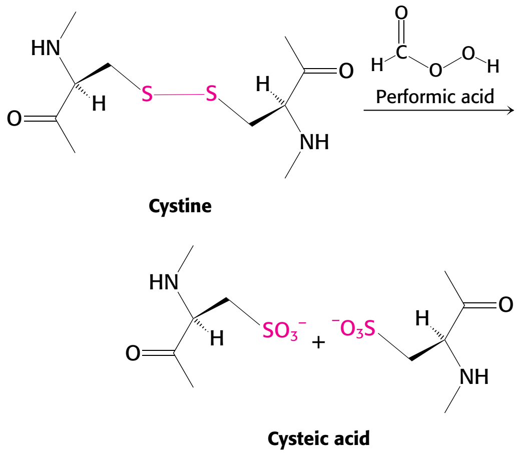
N-terminal analysis of (F,I,S) peptide: S

Cyanogen bromide treatment: (2S,F,G,I,K,L,M*,T,Y) (2S,L,W)

M*, methionine detected as homoserine

29. *Applications of two-dimensional electrophoresis.* Performic acid cleaves the disulfide linkage of cystine and converts the sulfhydryl groups into cysteic acid residues, which are then no longer capable of disulfide-bond formation.

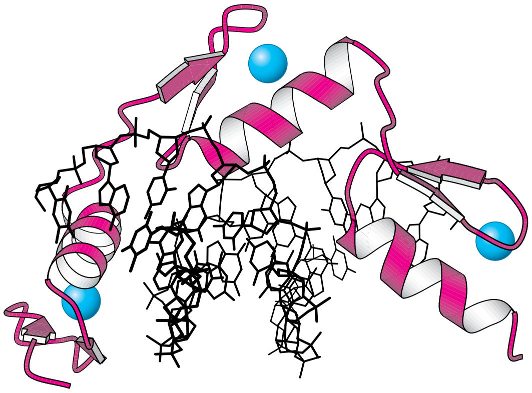
Consider the following experiment: You suspect that a protein containing three cysteine residues has a single disulfide bond. You digest the protein with trypsin and subject the mixture to electrophoresis along one end of a sheet of paper. After treating the paper with performic acid, you subject the sheet to electrophoresis in the perpendicular direction and stain it with a reagent that detects proteins. How would the paper appear if the protein did not contain any disulfide bonds? If the protein contained a single disulfide bond? Propose an experiment to identify which cysteine residues form the disulfide bond.



CHAPTER

4

DNA, RNA, and the Flow of Genetic Information



Family resemblance, very evident in this photograph of four sisters, results from having genes in common. Genes must be expressed to exert an effect, and proteins regulate such expression. One such regulatory protein, a zinc-finger protein (zinc ion is blue, protein is red), is shown bound to a control region of DNA (black). [(Left) © Nicholas Nixon, courtesy Fraenkel Gallery, San Francisco. (Right) Drawn from 1AAY.pdb.]

DNA and RNA are long linear polymers, called nucleic acids, that carry information in a form that can be passed from one generation to the next. These macromolecules consist of a large number of linked nucleotides, each composed of a sugar, a phosphate, and a base. Sugars linked by phosphates form a common backbone that plays a structural role, whereas the sequence of bases along a nucleic acid strand carries genetic information. The DNA molecule has the form of a *double helix*, a helical structure consisting of two complementary nucleic acid strands. Each strand serves as the template for the other in DNA replication. The genes of all cells and many viruses are made of DNA.

Genes specify the kinds of proteins that are made by cells, but DNA is not the direct template for protein synthesis. Rather, a DNA strand is copied into a class of RNA molecules called messenger RNA (mRNA), the information-carrying intermediates in protein synthesis. This process of transcription is followed by translation, the synthesis of proteins according to instructions given by mRNA templates. Information processing in all cells is quite complex. The scheme that underlies information processing at the level of gene expression was first proposed by Francis Crick in 1958.

$$\begin{tikzcd}
\text{DNA} \arrow[loop above, "\text{Replication}"] \arrow[r, "\text{Transcription}"] & \text{RNA} \arrow[r, "\text{Translation}"] & \text{Protein}
\end{tikzcd}$$

Crick called this scheme the *central dogma.* The basic tenets of this dogma are true, but, as we will see later, this scheme is not as simple as depicted.

OUTLINE

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

4.2 A Pair of Nucleic Acid Strands with Complementary Sequences Can Form a Double-Helical Structure

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons

This flow of information depends on the genetic code, which defines the relation between the sequence of bases in DNA (or its mRNA transcript) and the sequence of amino acids in a protein. The code is nearly the same in all organisms: a sequence of three bases, called a *codon,* specifies an amino acid. There is another step in the expression of most eukaryotic genes, which are mosaics of nucleic acid sequences called *introns* and *exons*. Both are transcribed, but before translation takes place, introns are cut out of newly synthesized RNA molecules, leaving mature RNA molecules with continuous exons. The existence of introns and exons has crucial implications for the evolution of proteins.

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

The nucleic acids DNA and RNA are well suited to function as the carriers of genetic information by virtue of their covalent structures. These macromolecules are *linear polymers* built up from similar units connected end to end (Figure 4.1). Each monomer unit within the polymer is a *nucleotide*. A single nucleotide unit consists of three components: a sugar, a phosphate, and one of four bases. *The sequence of bases in the polymer uniquely characterizes a nucleic acid and constitutes a form of linear information*—information analogous to the letters that spell a person's name.

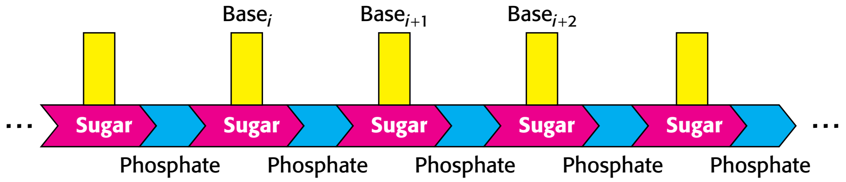
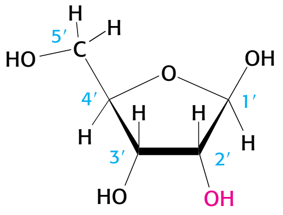


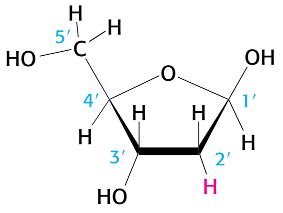
FIGURE 4.1 Polymeric structure of nucleic acids.

RNA and DNA differ in the sugar component and one of the bases

The sugar in *deoxyribonucleic acid* (DNA) is *deoxyribose*. The prefix deoxy indicates that the 2′-carbon atom of the sugar lacks the oxygen atom that is linked to the 2′-carbon atom of *ribose,* as shown in Figure 4.2. Note that sugar carbons are numbered with primes to differentiate them from atoms in the bases. The sugars in both nucleic acids are linked to one another by phosphodiester bridges. Specifically, the 3′-hydroxyl (3′-OH) group of the sugar moiety of one nucleotide is esterified to a phosphate group, which is, in turn, joined to the 5′-hydroxyl group of the adjacent sugar. The chain of sugars linked by phosphodiester bridges is referred to as the *backbone* of the nucleic acid (Figure 4.3). Whereas the backbone is constant in a nucleic

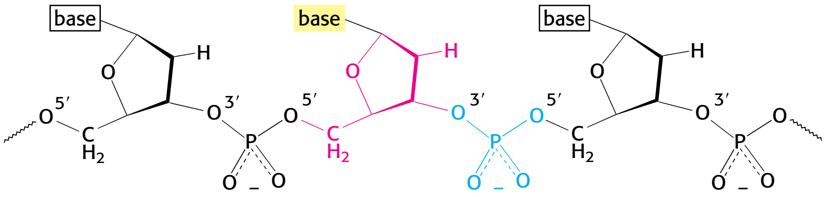


Ribose

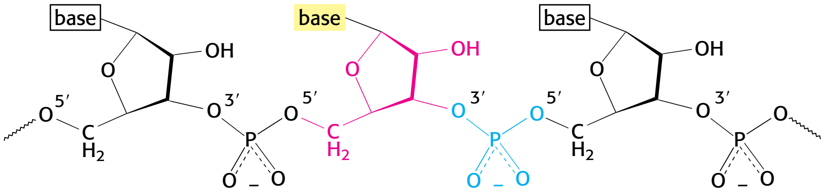


Deoxyribose

FIGURE 4.2 Ribose and deoxyribose. Atoms in sugar units are numbered with primes to distinguish them from atoms in bases (see Figure 4.4).



DNA



RNA

FIGURE 4.3 Backbones of DNA and RNA. The backbones of these nucleic acids are formed by 3′-to-5′ phosphodiester linkages. A sugar unit is highlighted in red and a phosphate group in blue.

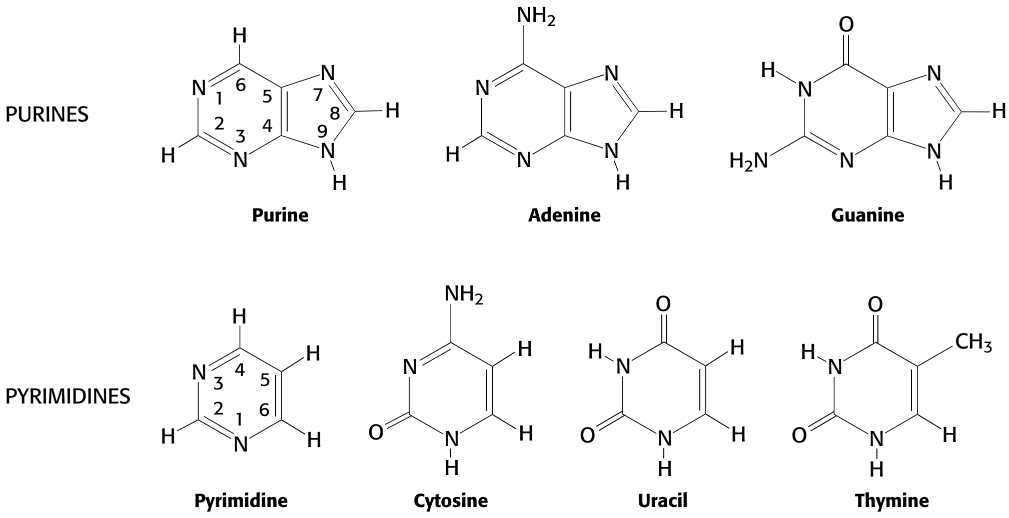


FIGURE 4.4 Purines and pyrimidines. Atoms within bases are numbered without primes. Uracil is present in RNA instead of thymine.

acid, the bases vary from one monomer to the next. Two of the bases of DNA are derivatives of *purine*—adenine (A) and guanine (G)—and two of *pyrimidine*—cytosine (C) and thymine (T), as shown in Figure 4.4.

Ribonucleic acid (RNA), like DNA, is a long unbranched polymer consisting of nucleotides joined by 3′-to-5′ phosphodiester linkages (Figure 4.3). The covalent structure of RNA differs from that of DNA in two respects. First, the sugar units in RNA are riboses rather than deoxyriboses. Ribose contains a 2′-hydroxyl group not present in deoxyribose. Second, one of the four major bases in RNA is uracil (U) instead of thymine (T).

Note that each phosphodiester bridge has a negative charge. This negative charge repels nucleophilic species such as hydroxide ions, which are capable of hydrolytic attack on the phosphate backbone. This resistance is crucial for maintaining the integrity of information stored in nucleic acids. The absence of the 2′-hydroxyl group in DNA further increases its resistance to hydrolysis. The greater stability of DNA probably accounts for its use rather than RNA as the hereditary material in all modern cells and in many viruses.

Nucleotides are the monomeric units of nucleic acids

The building blocks of nucleic acids and the precursors of these building blocks play many other roles throughout the cell—for instance, as energy currency and as molecular signals. Consequently, it is important to be familiar with the nomenclature of nucleotides and their precursors. A unit consisting of a base bonded to a sugar is referred to as a *nucleoside*. The four nucleoside units in RNA are called *adenosine, guanosine, cytidine*, and *uridine*, whereas those in DNA are called *deoxyadenosine, deoxyguanosine, deoxycytidine*, and *thymidine*. In each case, N-9 of a purine or N-1 of a pyrimidine is attached to C-1′ of the sugar by an *N*-glycosidic linkage (Figure 4.5). The base lies above the plane of the sugar when the structure is written in the standard orientation; that is, the configuration of the *N*-glycosidic linkage is β (Section 11.1). Note that thymidine contains deoxyribose; by convention, the prefix deoxy is not added because thymine-containing nucleosides are only rarely found in RNA.

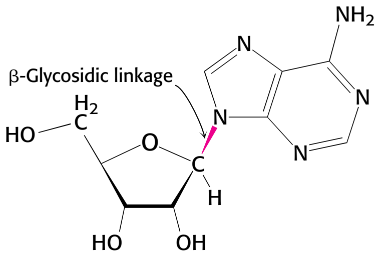


FIGURE 4.5 β-Glycosidic linkage in a nucleoside.

A *nucleotide* is a nucleoside joined to one or more phosphoryl groups by an ester linkage. *Nucleotide triphosphates, nucleosides joined to three phosphoryl groups, are the monomers—the building blocks—that are linked to form RNA and DNA.* The four nucleotide units that link to form DNA are nucleotide monophosphates called *deoxyadenylate, deoxyguanylate, deoxycytidylate*, and *thymidylate*. Similarly, the most common nucleotides that link to form RNA are nucleotide monophosphates adenylate, guanylate, cytidylate and uridylate.

This nomenclature does not describe the number of phosphoryl groups or the site of attachment to carbon of the ribose. A more precise nomenclature is also commonly used. A compound formed by the attachment of a phosphoryl group to C-5′ of a nucleoside sugar (the most common site of phosphate esterification) is called a *nucleoside 5′-phosphate* or a *5′-nucleotide*. In this naming system for nucleotides, the number of phosphoryl groups and the attachment site are designated. Look, for example, at *adenosine 5′-triphosphate* (ATP; Figure 4.6). This nucleotide is tremendously important because, in addition to being a building block for RNA, it is the most commonly used energy currency. The energy released from cleavage of the triphosphate group is used to power many cellular processes (Chapter 15). Another nucleotide is deoxyguanosine 3′-monophosphate (3′-dGMP; Figure 4.6). This nucleotide differs from ATP in that it contains guanine rather than adenine, deoxyribose rather than ribose (indicated by the prefix "d"), and one rather than three phosphoryl groups. In addition, the phosphoryl group is esterified to the hydroxyl group in the 3′ rather than the 5′ position.

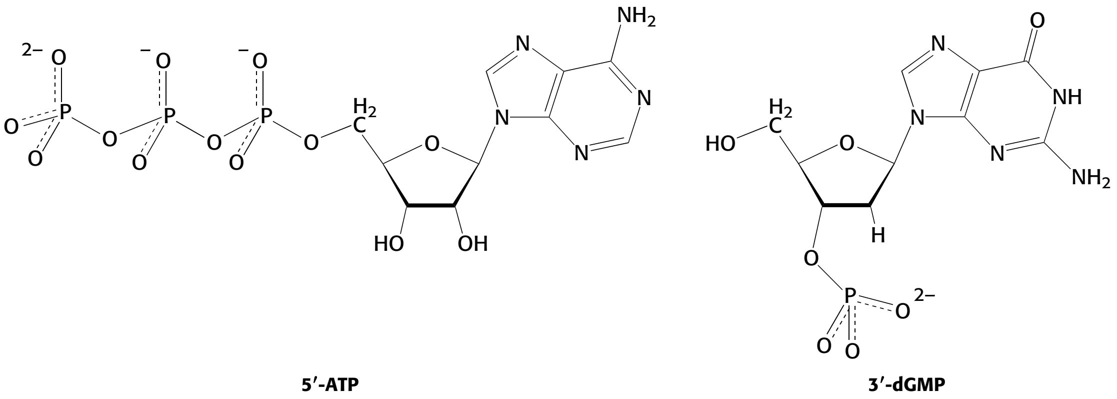


FIGURE 4.6 Nucleotides adenosine 5′-triphosphate (5′-ATP) and deoxyguanosine 3′-monophosphate (3′-dGMP).

DNA molecules are very long and have directionality

Scientific communication frequently requires the sequence of a nucleic acid—in some cases, a sequence thousands of nucleotides in length—to be written like that on page 17. Rather than writing the cumbersome chemical structures, scientists have adopted the use of abbreviations. The abbreviated notations pApCpG or ACG denote a trinucleotide of DNA consisting of the building blocks deoxyadenylate monophosphate, deoxycytidylate monophosphate, and deoxyguanylate monophosphate linked by a phosphodiester bridge, where "p" denotes a phosphoryl group (Figure 4.7). The 5′ end will often have a phosphoryl group attached to the 5′-OH group. Note that, like a polypeptide (Section 2.2), *a DNA chain has directionality,* commonly called polarity. One end of the chain has a free 5′-OH group (or a 5′-OH group attached to a phosphoryl group) and the other end has a free 3′-OH group, neither of which is linked to another nucleotide. By convention, *the base sequence is written in the 5′-to-3′ direction*. Thus, ACG indicates that the unlinked 5′-OH group is on deoxyadenylate, whereas the unlinked 3′-OH group is on deoxyguanylate. Because of this polarity, ACG and GCA correspond to different compounds.

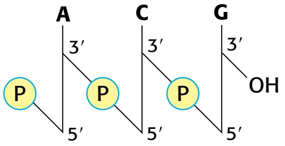


FIGURE 4.7 Structure of a DNA strand. The strand has a 5′ end, which is usually attached to a phosphoryl group, and a 3′ end, which is usually a free hydroxyl group.

A striking characteristic of naturally occurring DNA molecules is their length. A DNA molecule must comprise many nucleotides to carry the genetic information necessary for even the simplest organisms. For example, the DNA of a virus such as polyoma, which can cause cancer in certain organisms, consists of two paired strands of DNA, each 5100 nucleotides in length. The *E. coli* genome is a single DNA molecule consisting of two strands of 4.6 million nucleotides each (Figure 4.8).

The DNA molecules of higher organisms can be much larger. The human genome comprises approximately 3 billion nucleotides in each strand of DNA, divided among 24 distinct molecules of DNA called chromosomes (22 autosomal chromosomes plus the X and Y sex chromosomes) of different sizes. One of the largest known DNA molecules is found in the Indian muntjac, an Asiatic deer; its genome is nearly as large as the human genome but is distributed on only 3 chromosomes (Figure 4.9). The largest of these chromosomes has two strands of more than 1 billion nucleotides each. If such a DNA molecule could be fully extended, it would stretch more than 1 foot in length. Some plants contain even larger DNA molecules.

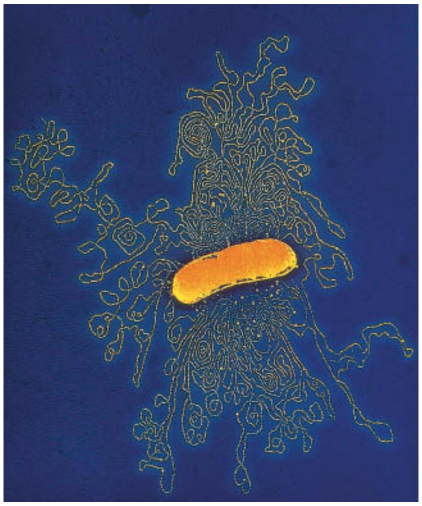


FIGURE 4.8 Electron micrograph of part of the *E. coli* genome. [Dr. Gopal Murti/Science Photo Library/Photo Researchers.]

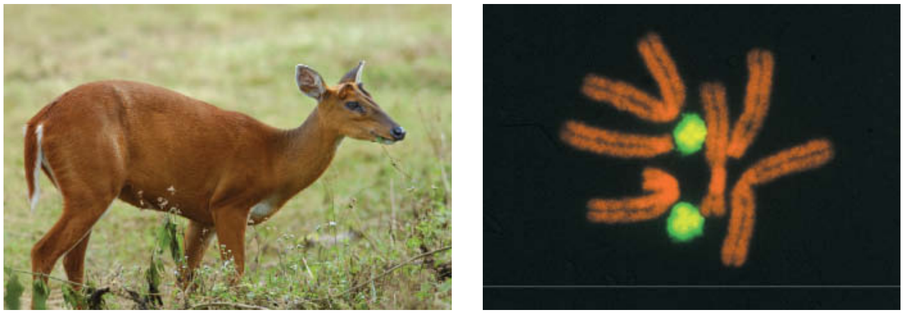


FIGURE 4.9 The Indian muntjac and its chromosomes. Cells from a female Indian muntjac (right) contain three pairs of very large chromosomes (stained orange). The cell shown is a hybrid containing a pair of human chromosomes (stained green) for comparison. [(Left) Hugh Lansdown/Shutterstock. (Right) J.–Y. Lee, M. Koi, E. J. Stanbridge, M. Oshimura, A. T. Kumamoto, and A. P. Feinberg. *Nat. Genet.* 7:30, 1994.]

4.2 A Pair of Nucleic Acid Strands with Complementary Sequences Can Form a Double-Helical Structure

As discussed in Chapter 1, the covalent structure of nucleic acids accounts for their ability to carry information in the form of a sequence of bases along a nucleic acid strand. The bases on the two separate nucleic acid strands form *specific base pairs* in such a way that a helical structure is formed. The double-helical structure of DNA facilitates the *replication* of the genetic material—that is, the generation of two copies of a nucleic acid from one.

The double helix is stabilized by hydrogen bonds and van der Waals interactions

The ability of nucleic acids to form specific base pairs was discovered in the course of studies directed at determining the three-dimensional structure of DNA. Maurice Wilkins and Rosalind Franklin obtained x-ray diffraction photographs of fibers of DNA (Figure 4.10). The characteristics of these diffraction patterns indicated that DNA is formed of two strands that wind in a regular helical structure. From these data and others, James Watson and Francis Crick deduced a structural model for DNA that accounted for the diffraction pattern

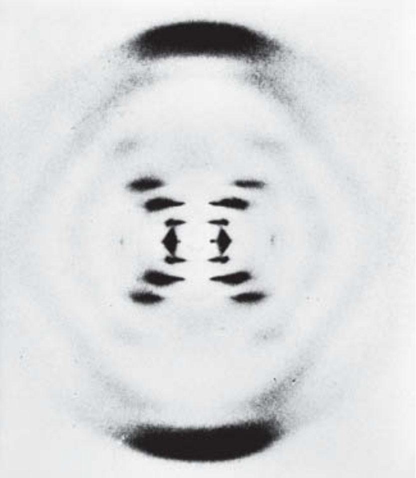


FIGURE 4.10 X-ray diffraction photograph of a hydrated DNA fiber. When crystals of a biomolecule are irradiated with x-rays, the x-rays are diffracted and these diffracted x-rays are seen as a series of spots, called reflections, on a screen behind the crystal. The structure of the molecule can be determined by the pattern of the reflections (Section 3.5). In regard to DNA crystals, the central cross is diagnostic of a helical structure. The strong arcs on the meridian arise from the stack of nucleotide bases, which are 3.4 Å apart. [Science Photo Library.]

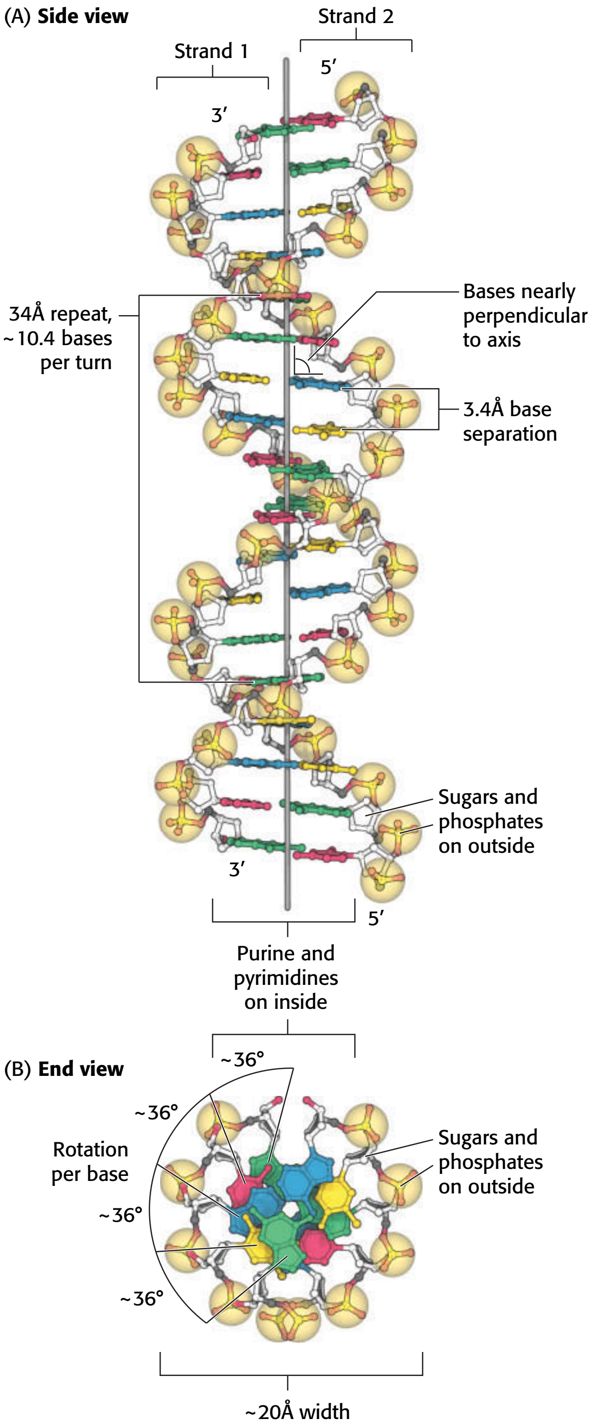

(A) Side view
Strand 1
Strand 2
5′
3′
34Å repeat, ~10.4 bases per turn
Bases nearly perpendicular to axis
3.4Å base separation
Sugars and phosphates on outside
3′
5′
Purine and pyrimidines on inside
(B) End view
~36°
~36°
Rotation per base
~36°
~36°
Sugars and phosphates on outside
~20Å width

FIGURE 4.11 Watson–Crick model of double-helical DNA. (A) Side view. Adjacent bases are separated by 3.4 Å. The structure repeats along the helical axis (vertical) at intervals of 34 Å, which corresponds to approximately 10 nucleotides on each chain. (B) Axial view, looking down the helix axis, reveals a rotation of 36^{o} per base and shows that the bases are stacked on top of one another [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry*: *A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 33.11.].

and was the source of some remarkable insights into the functional properties of nucleic acids (Figure 4.11).

The features of the Watson–Crick model of DNA deduced from the diffraction patterns are:

1. Two helical polynucleotide strands are coiled around a common axis with a right-handed screw sense (pp. 40-41). The strands are antiparallel, meaning that they have opposite directionality.
2. The sugar–phosphate backbones are on the outside and the purine and pyrimidine bases lie on the inside of the helix.
3. The bases are nearly perpendicular to the helix axis, and adjacent bases are separated by approximately 3.4 Å. The helical structure repeats on the order of every 34 Å, with about 10.4 bases per turn of helix. There is a rotation of nearly 36 degrees per base (360 degrees per full turn/10.4 bases per turn).
4. The diameter of the helix is about 20 Å.

How is such a regular structure able to accommodate an arbitrary sequence of bases, given the different sizes and shapes of the purines and pyrimidines? In attempting to answer this question, Watson and Crick discovered that guanine can be paired with cytosine and adenine with thymine to form base pairs that have essentially the same shape (Figure 4.12). These base pairs are held together by specific hydrogen bonds, which, although weak (4–21 kJ mol^{-1}, or 1–5 kcal mol^{-1}), stabilize the helix because of their large numbers in a DNA molecule. These *base-pairing rules* account for the observation, originally made by Erwin Chargaff in 1950, that the ratios of adenine to thymine and of guanine to cytosine are nearly the same in all species studied, whereas the adenine-to-guanine ratio varies considerably (Table 4.1).

Inside the helix, the bases are essentially stacked one on top of another (Figure 4.11B). The stacking of base pairs contributes to the stability of the double helix in two ways. First, the formation of the double helix is facilitated by the hydrophobic effect (p. 9). The hydrophobic bases cluster in the interior of the helix away

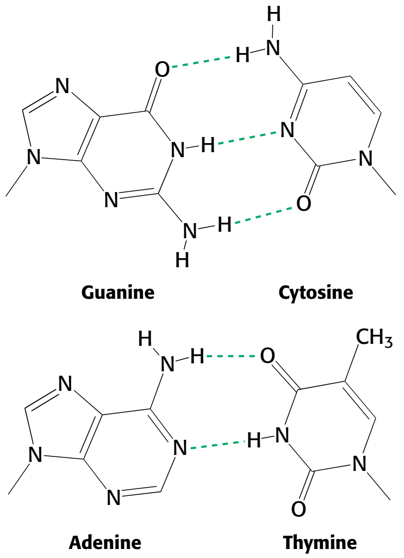

Guanine
Cytosine
Adenine
Thymine
CH3

FIGURE 4.12 Structures of the base pairs proposed by Watson and Crick.

TABLE 4.1 Base compositions experimentally determined for a variety of organisms

Organism	A : T	G : C	A : G
Human being	1.00	1.00	1.56
Salmon	1.02	1.02	1.43
Wheat	1.00	0.97	1.22
Yeast	1.03	1.02	1.67
Escherichia coli	1.09	0.99	1.05
Serratia marcescens	0.95	0.86	0.70

from the surrounding water, whereas the more polar surfaces are exposed to water. This arrangement is reminiscent of protein folding, where hydrophobic amino acids are in the protein's interior and the hydrophilic amino acids are on the exterior (Section 2.4). Second, the stacked base pairs attract one another through van der Waals forces (p. 8), appropriately referred to as *stacking forces,* further contributing to stabilization of the helix (Figure 4.13). The energy associated with a single van der Waals interaction is quite small, typically from 2 to 4 kJ mol^{-1} (0.5–1.0 kcal mol^{-1}). In the double helix, however, a large number of atoms are in van der Waals contact, and the net effect, summed over these atom pairs, is substantial. In addition, base stacking in DNA is favored by the conformations of the somewhat rigid five-membered rings of the backbone sugars.

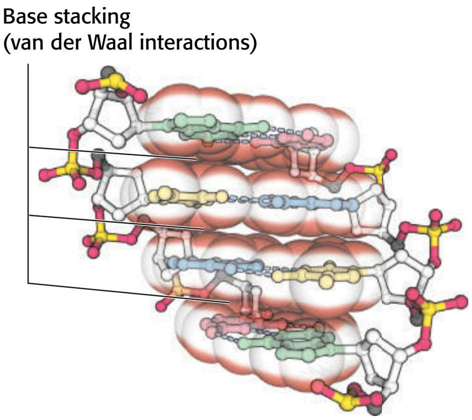


FIGURE 4.13 A side view of DNA. Base pairs are stacked nearly one on top of another in the double helix. The stacked bases interact with van der Waals forces. Such stacking forces help stabilize the double helix. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 33.13.].

DNA can assume a variety of structural forms

Watson and Crick based their model (known as the *B-DNA helix*) on x-ray diffraction patterns of highly hydrated DNA fibers, which provided information about properties of the double helix that are averaged over its constituent residues. Under physiological conditions, most DNA is in the B form. X-ray diffraction studies of less-hydrated DNA fibers revealed a different form called *A-DNA*. Like B-DNA, A-DNA is a right-handed double helix made up of anti-parallel strands held together by Watson–Crick base-pairing. The A-form helix is wider and shorter than the B-form helix, and its base pairs are tilted rather than perpendicular to the helix axis (Figure 4.14).

If the A-form helix were simply a property of dehydrated DNA, it would be of little significance. However, double-stranded regions of RNA and at least some RNA–DNA hybrids adopt a double-helical form very similar to that of A-DNA. What is the biochemical basis for differences between the two forms of DNA? Many of the structural differences between B-DNA and A-DNA arise from different puckerings of their ribose units (Figure 4.15). In A-DNA, C-3′ lies out of the plane (a conformation referred to as C-3′ endo) formed by the other four atoms of the ring; in B-DNA, C-2′ lies out of the plane (a conformation called C-2′ endo). The C-3′-endo puckering in A-DNA leads to an 11-degree tilting of

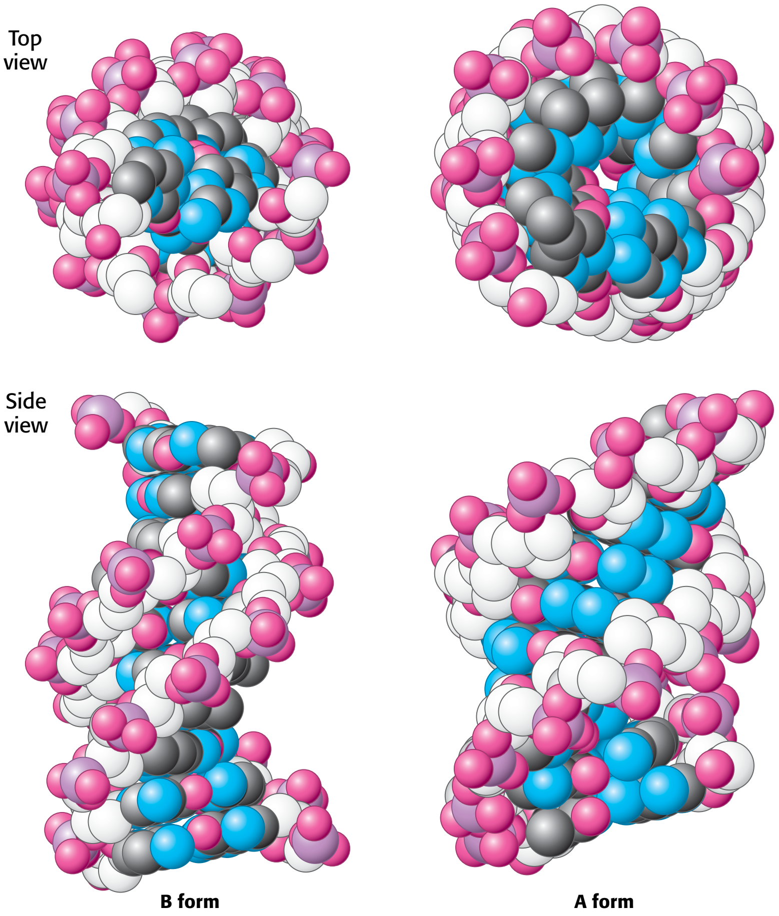


FIGURE 4.14 B-form and A-form DNA. Space-filling models of 10 base pairs of B-form and A-form DNA depict their right-handed helical structures. *Notice* that the B-form helix is longer and narrower than the A-form helix. The carbon atoms of the backbone are shown in white. [Drawn from 1BNA.pdb and 1DNZ.pdb.]

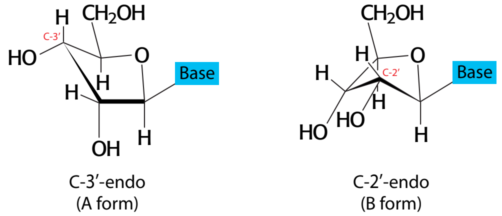


FIGURE 4.15 Sugar pucker. In A-form DNA, the C-3′ carbon atom lies above the approximate plane defined by the four other sugar nonhydrogen atoms (called C-3′ endo). In B-form DNA, each deoxyribose is in a C-2′-endo conformation, in which C-2′ lies out of the plane.

the base pairs away from perpendicular to the helix. RNA helices are further induced to take the A-DNA form because of steric hindrance from the 2′-hydroxyl group: the 2′-oxygen atom would be too close to three atoms of the adjoining phosphoryl group and to one atom in the next base. In an A-form helix, in contrast, the 2′-oxygen atom projects outward, away from other atoms. The phosphoryl and other groups in the A-form helix bind fewer H_2O molecules than do those in B-DNA. Hence, dehydration favors the A form.

Z-DNA is a left-handed double helix in which backbone phosphates zigzag

A third type of double helix is *left-handed,* in contrast with the *right-handed* screw sense of the A and B helices. Furthermore, the phosphoryl groups in the backbone are *zigzagged;* hence, this form of DNA is called *Z-DNA* (Figure 4.16).

Although the biological role of Z-DNA is still under investigation, Z-DNA-binding proteins have been isolated, one of which is required for viral pathogenesis of poxviruses, including variola, the agent of smallpox. The existence of Z-DNA shows that DNA is a flexible, dynamic molecule whose parameters are not as fixed as depictions suggest. The properties of A-, B-, and Z-DNA are compared in Table 4.2.

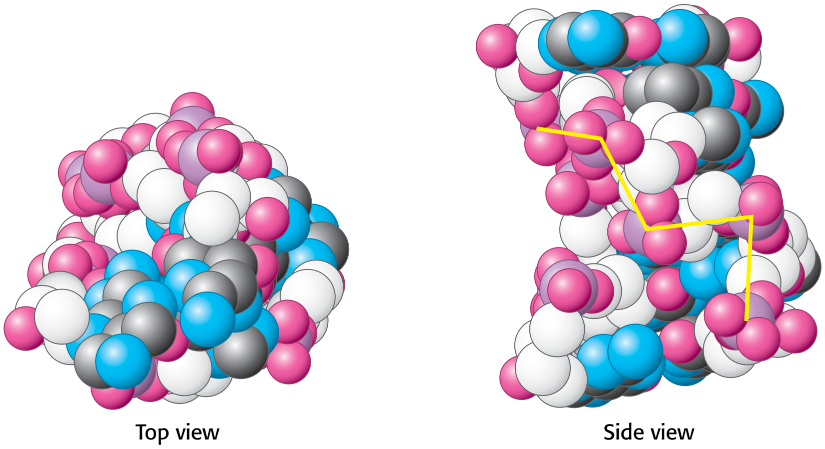


FIGURE 4.16 Z-DNA. DNA oligomers such as CGCGCG adopt an alternative conformation under some conditions. This conformation is called Z-DNA because the phosphoryl groups zigzag along the backbone. [Drawn from 131D.pdb.]

TABLE 4.2 Comparison of A-, B-, and Z-DNA

	A	B	Z
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	~26 Å	~20 Å	~18 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond*	*anti*	*anti*	Alternating *anti* and *syn*
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from perpendicular to helix axis	19 degrees	1 degree	9 degrees

**Syn* and *anti* refer to the orientation of the *N*-glycosidic bond between the base and deoxyribose. In the *anti* orientation, the base extends away from the deoxyribose. In the *syn* orientation, the base is above the deoxyribose. Pyrimidines can be in *anti* orientations only, whereas purines can be *anti* or *syn*.

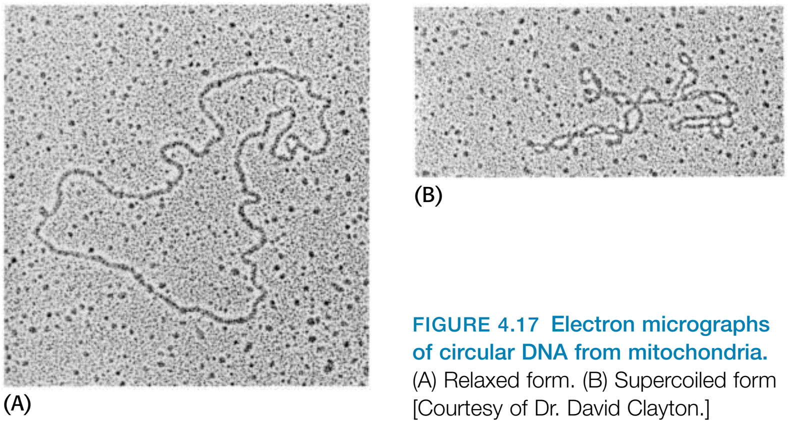


FIGURE 4.17 Electron micrographs of circular DNA from mitochondria. (A) Relaxed form. (B) Supercoiled form [Courtesy of Dr. David Clayton.]

Some DNA molecules are circular and supercoiled

The DNA molecules in human chromosomes are linear. However, electron microscopic and other studies have shown that intact DNA molecules from bacteria and archaea are circular (Figure 4.17A). The term *circular* refers to the continuity of the DNA strands, not to their geometric form. DNA molecules inside cells necessarily have a very compact shape. Note that the *E. coli* chromosome, fully extended, would be about 1000 times as long as the greatest diameter of the bacterium.

A closed DNA molecule has a property unique to circular DNA. The axis of the double helix can itself be twisted or supercoiled into a *superhelix* (Figure 4.17B). A circular DNA molecule without any superhelical turns is known as a *relaxed molecule*. Supercoiling is biologically important for two reasons. First, *a supercoiled DNA molecule is more compact than its relaxed counterpart*. Second, *supercoiling may hinder or favor the capacity of the double helix to unwind and thereby affect the interactions between DNA and other molecules.* These topological features of DNA will be considered further in Chapter 28.

Single-stranded nucleic acids can adopt elaborate structures

Single-stranded nucleic acids often fold back on themselves to form well-defined structures. Such structures are especially prominent in RNA and RNA-containing complexes such as the ribosome—a large complex of RNAs and proteins on which proteins are synthesized.

The simplest and most-common structural motif formed is a *stem-loop,* created when two complementary sequences within a single strand come together to form double-helical structures (Figure 4.18). In many cases, these double helices are made up entirely of Watson–Crick base pairs. In other cases, however, the structures include mismatched base pairs or unmatched bases that bulge out from the helix. Such mismatches destabilize

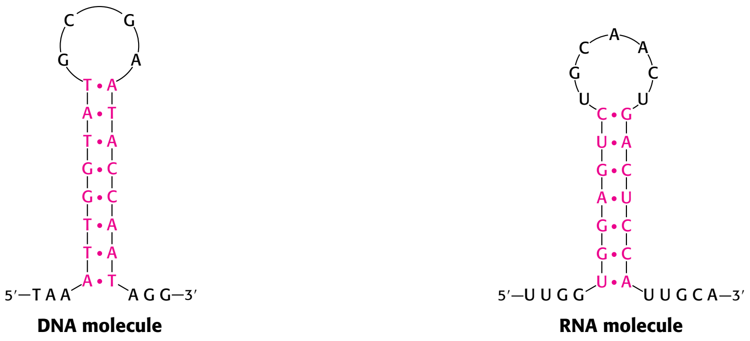


FIGURE 4.18 Stem-loop structures. Stem-loop structures can be formed from single-stranded DNA and RNA molecules.

the local structure but introduce deviations from the standard double-helical structure that can be important for higher-order folding and for function (Figure 4.19).

Single-stranded nucleic acids can adopt structures that are more complex than simple stem-loops through the interaction of more widely separated bases. Often, three or more bases interact to stabilize these structures. In such cases, hydrogen-bond donors and acceptors that do not participate in Watson–Crick base pairs participate in hydrogen bonds to form nonstandard pairings. Metal ions such as magnesium ion (Mg^{2+}) often assist in the stabilization of these more elaborate structures. These complex structures allow RNA to perform a host of functions that the double-stranded DNA molecule cannot. Indeed, the complexity of some RNA molecules rivals that of proteins, and these RNA molecules perform a number of functions that had formerly been thought the exclusive domain of proteins.

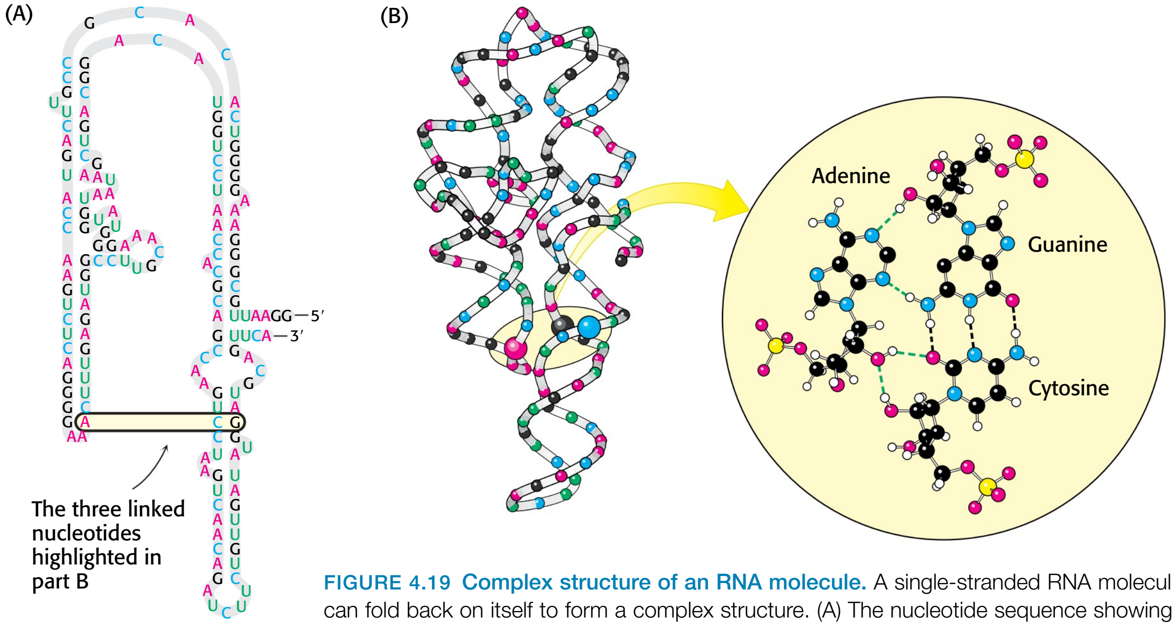


FIGURE 4.19 Complex structure of an RNA molecule. A single-stranded RNA molecule can fold back on itself to form a complex structure. (A) The nucleotide sequence showing Watson–Crick base pairs and other nonstandard base pairings in stem-loop structures. (B) The three-dimensional structure and one important long-range interaction between three bases. In the three-dimensional structure at the left, cytidine nucleotides are shown in blue, adenosine in red, guanosine in black, and uridine in green. In the detailed projection, hydrogen bonds within the Watson–Crick base pair are shown as dashed black lines; additional hydrogen bonds are shown as dashed green lines.

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

The double-helical model of DNA and the presence of specific base pairs immediately suggested how the genetic material might replicate. *The sequence of bases of one strand of the double helix precisely determines the sequence of the other strand:* a guanine base on one strand is always paired with a cytosine base on the other strand, and so on. Thus, separation of a double helix into its two component strands would yield two single-stranded templates onto which new double helices could be constructed, each of which would have the same sequence of bases as the parent double helix. Consequently, as DNA is replicated, one of the strands of each daughter DNA molecule is newly synthesized, whereas the other is passed unchanged from the parent DNA molecule. This distribution of parental atoms is achieved by *semiconservative replication.*

Differences in DNA density established the validity of the semiconservative replication hypothesis

Matthew Meselson and Franklin Stahl carried out a critical test of this hypothesis in 1958. They labeled the parent DNA with ^{15}N, a heavy isotope of nitrogen, to make it denser than ordinary DNA. The labeled DNA was generated by growing *E. coli* for many generations in a medium that contained $^{15}NH_4Cl$ as the sole nitrogen source. After the incorporation of heavy nitrogen was complete, the bacteria were abruptly transferred to a medium that contained ^{14}N, the ordinary isotope of nitrogen. The question asked was: What is the distribution of ^{14}N and ^{15}N in the DNA molecules after successive rounds of replication?

The distribution of ^{14}N and ^{15}N was revealed by the technique of *density-gradient equilibrium sedimentation*. A small amount of DNA was dissolved in a concentrated solution of cesium chloride having a density close to that of the DNA (1.7 g cm^{-3}). This solution was centrifuged until it was nearly at equilibrium. At that point, the opposing processes of sedimentation and diffusion created a gradient in the concentration of cesium chloride across the centrifuge cell. The result was a stable density gradient ranging from 1.66 to 1.76 g cm^{-3}. The DNA molecules in this density gradient were driven by centrifugal force into the region where the solution's density was equal to their own. The DNA yielded a narrow band that was detected by its absorption of ultraviolet light. A mixture of ^{14}N DNA and ^{15}N DNA molecules gave clearly separate bands because they differ in density by about 1% (Figure 4.20).

DNA was extracted from the bacteria at various times after they were transferred from a ^{15}N to a ^{14}N medium. Analysis of these samples by the density-gradient technique showed that there was a single band of DNA after one generation. The density of this band was precisely halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands (Figure 4.21).

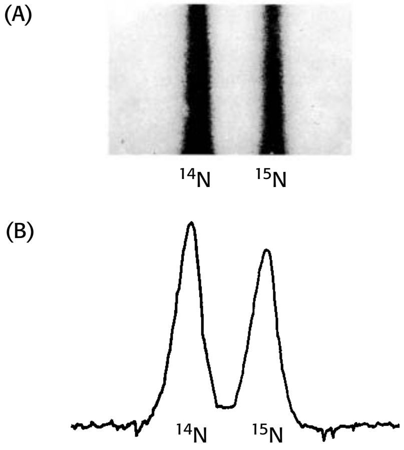


FIGURE 4.20 Resolution of ^{14}N DNA and ^{15}N DNA by density-gradient centrifugation. (A) Ultraviolet-absorption photograph of a centrifuged cell showing the two distinct bands of DNA. (B) Densitometric tracing of the absorption photograph. [Data from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

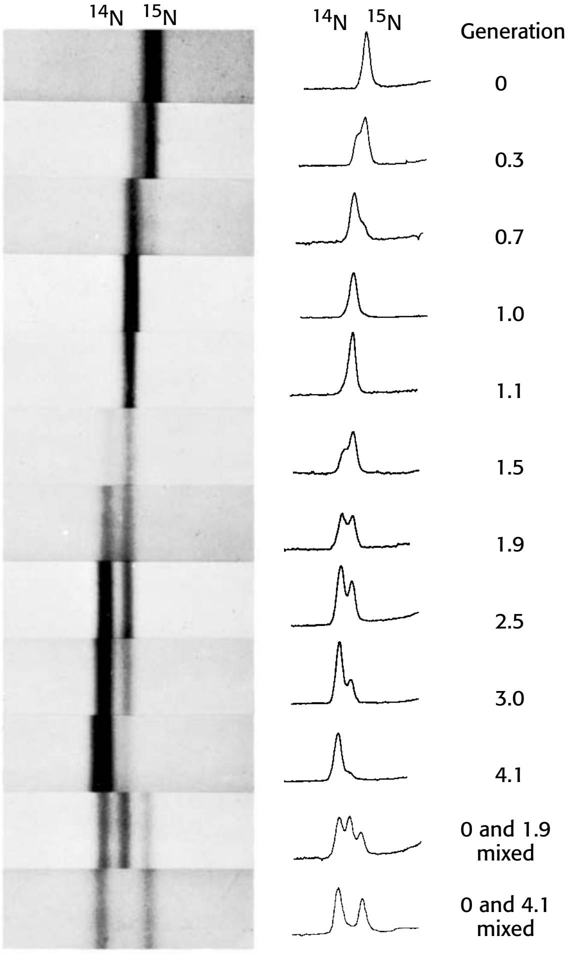


FIGURE 4.21 Detection of semiconservative replication of *E. coli* DNA by density-gradient centrifugation. The position of a band of DNA depends on its content of ^{14}N and ^{15}N. After 1.0 generation, all of the DNA molecules were hybrids containing equal amounts of ^{14}N and ^{15}N. [Data from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

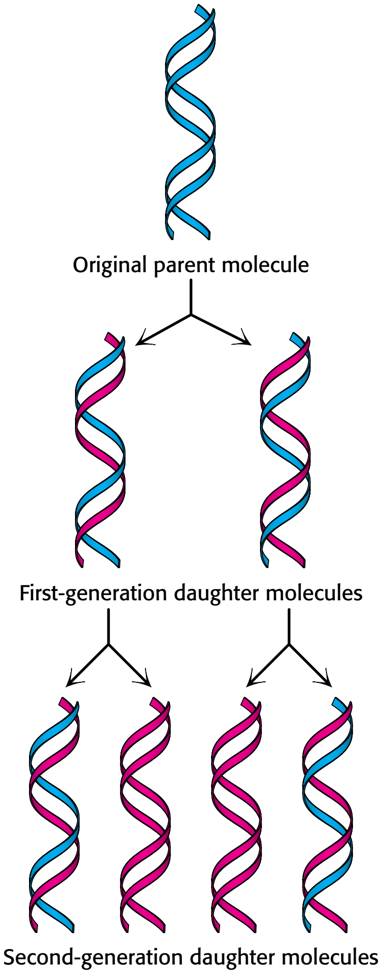

Original parent molecule
First-generation daughter molecules
Second-generation daughter molecules

FIGURE 4.22 Diagram of semiconservative replication. Parental DNA is shown in blue and newly synthesized DNA in red. [Information from M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

The absence of ^{15}N DNA indicated that parental DNA was not preserved as an intact unit after replication. The absence of ^{14}N DNA indicated that all the daughter DNA derived some of their atoms from the parent DNA. This proportion had to be half because the density of the hybrid DNA band was halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands.

After two generations, there were equal amounts of two bands of DNA. One was hybrid DNA, and the other was ^{14}N DNA. Meselson and Stahl concluded from these incisive experiments that replication was semiconservative, and so each new double helix contains a parent strand and a newly synthesized strand. Their results agreed perfectly with the Watson–Crick model for DNA replication (Figure 4.22).

The double helix can be reversibly melted

During DNA replication and transcription, the two strands of the double helix must be separated from each other, at least in a local region. The two strands of a DNA helix readily come apart when the hydrogen bonds between base pairs are disrupted. In the laboratory, the double helix can be disrupted by heating a solution of DNA or by adding acid or alkali to ionize its bases. The dissociation of the double helix is called *melting* because it occurs abruptly at a certain temperature. The *melting temperature* (T_m) of DNA is defined as the temperature at which half the helical structure is lost. Inside cells, however, the double helix is not melted by the addition of heat. Instead, proteins called *helicases* use chemical energy (from ATP) to disrupt the helix (Chapter 28).

Stacked bases in nucleic acids absorb less ultraviolet light than do unstacked bases, an effect called *hypochromism*. Thus, the melting of nucleic acids is readily monitored by measuring their absorption of light, which is maximal at a wavelength of 260 nm (Figure 4.23).

Separated complementary strands of nucleic acids spontaneously reassociate to form a double helix when the temperature is lowered below T_m. This renaturation process is sometimes called *annealing*. The facility with which double helices can be melted and then reassociated is crucial for the biological functions of nucleic acids.

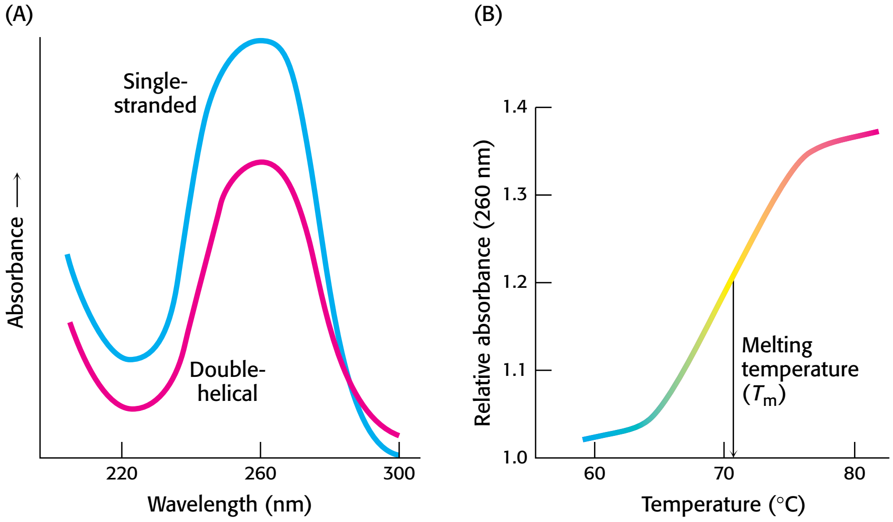


FIGURE 4.23 Hypochromism. (A) Single-stranded DNA absorbs light more effectively than does double-helical DNA. (B) The absorbance of a DNA solution at a wavelength of 260 nm increases when the double helix is melted into single strands.

The ability to melt and reanneal DNA reversibly in the laboratory provides a powerful tool for investigating sequence similarity. For instance, DNA molecules from two different organisms can be melted and allowed to reanneal, or *hybridize,* in the presence of each other. If the sequences are similar, hybrid DNA duplexes, with DNA from each organism contributing a strand of the double helix, can form. The degree of hybridization is an indication of the relatedness of the genomes and hence the organisms. Similar hybridization experiments with RNA and DNA can locate genes in a cell's DNA that correspond to a particular RNA. We will return to this important technique in Chapter 5.

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

We now turn to the molecular mechanism of DNA replication. The full replication machinery in a cell comprises more than 20 proteins engaged in intricate and coordinated interplay. In 1958, Arthur Kornberg and his colleagues isolated from *E. coli* the first known of the enzymes, called *DNA polymerases,* that promote the formation of the bonds joining units of the DNA backbone. *E. coli* has a number of DNA polymerases, designated by roman numerals, that participate in DNA replication and repair (Chapter 28).

DNA polymerase catalyzes phosphodiester-bridge formation

DNA polymerases catalyze the step-by-step addition of deoxyribonucleotide units to a DNA strand (Figure 4.24). The reaction catalyzed, in its simplest form, is

$$(\mathrm{DNA})_n + \mathrm{dNTP} \rightleftharpoons (\mathrm{DNA})_{n+1} + \mathrm{PP_i}$$

where dNTP stands for any deoxyribonucleotide and $\mathrm{PP_i}$ is a pyrophosphate ion.

DNA synthesis has the following characteristics:

1. The reaction requires all four activated precursors—that is, *the deoxynucleoside 5′-triphosphates dATP, dGTP, dCTP, and TTP*—as well as Mg^{2+} ion.

2. *The new DNA strand is assembled directly on a preexisting DNA template.* DNA polymerases catalyze the formation of a phosphodiester linkage efficiently only if the base on the incoming nucleoside triphosphate is complementary to the base on the template strand. Thus, DNA polymerase is a *template-directed enzyme* that synthesizes a product with a base sequence complementary to that of the template.

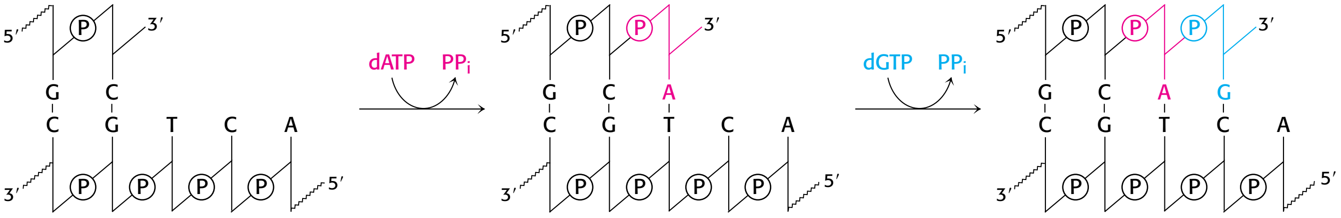


FIGURE 4.24 Polymerization reaction catalyzed by DNA polymerases.

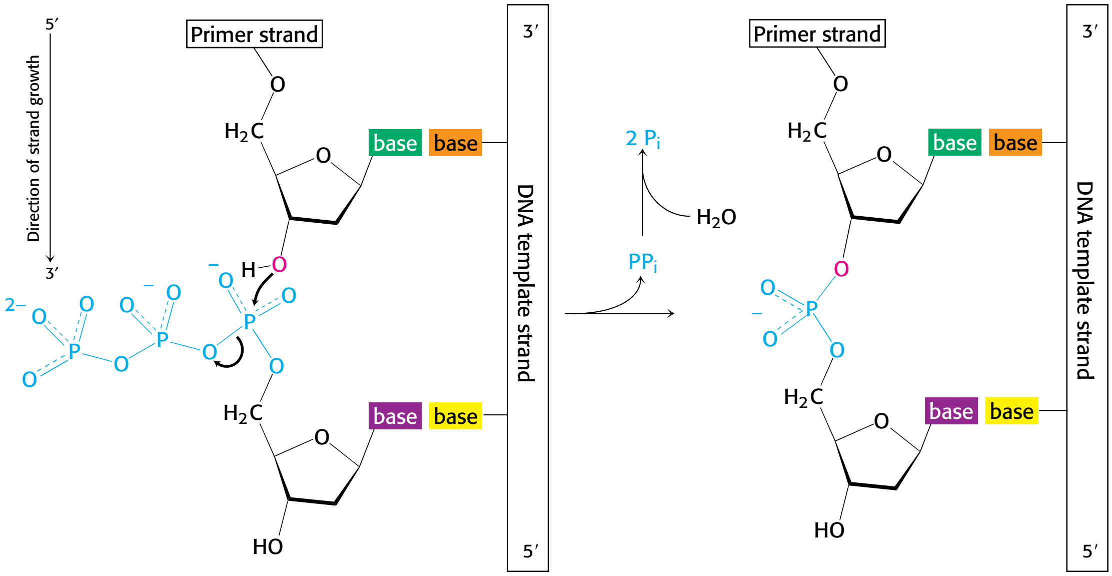


FIGURE 4.25 Strand-elongation reaction. DNA polymerases catalyze the formation of a phosphodiester bridge. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 34.2.]

3. *DNA polymerases require a primer to begin synthesis*. A *primer* strand having a free 3′-OH group must be already bound to the template strand. The chain-elongation reaction catalyzed by DNA polymerases is a nucleophilic attack by the 3′-OH terminus of the growing strand on the innermost phosphorus atom of the deoxynucleoside triphosphate (Figure 4.25). A phosphodiester bridge is formed and pyrophosphate is released. The subsequent hydrolysis of pyrophosphate to yield two ions of orthophosphate (P_i) by *pyrophosphatase* helps drive the polymerization forward. *Elongation of the DNA chain proceeds in the 5′-to-3′ direction.*

4. *Many DNA polymerases are able to correct mistakes in DNA by removing mismatched nucleotides.* These polymerases have a distinct nuclease activity that allows them to excise incorrect bases by a separate reaction. This nuclease activity contributes to the remarkably high fidelity of DNA replication, which has an error rate of less than 10^{-8} per base pair.

The genes of some viruses are made of RNA

Genes in all cellular organisms are made of DNA. The same is true for some viruses but, for others, the genetic material is RNA. Viruses are genetic elements enclosed in protein coats that can move from one cell to another but are not capable of independent growth. A well-studied example of an RNA virus is the tobacco mosaic virus, which infects the leaves of tobacco plants. This virus consists of a single strand of RNA (6390 nucleotides) surrounded by a protein coat of 2130 identical subunits. An RNA polymerase that takes direction from an RNA template, called an *RNA-directed RNA polymerase*, copies the viral RNA. The infected cells die because of virus-instigated programmed cell death; in essence, the virus instructs the cell to commit suicide. Cell death results in discoloration in the tobacco leaf in a variegated pattern, hence the name mosaic virus.

Another important class of RNA virus comprises the *retroviruses*, so called because the genetic information flows from RNA to DNA rather than from DNA to RNA. This class includes human immunodeficiency virus 1 (HIV-1), the cause of acquired immunodeficiency syndrome (AIDS), as well as a number of RNA viruses that produce tumors in

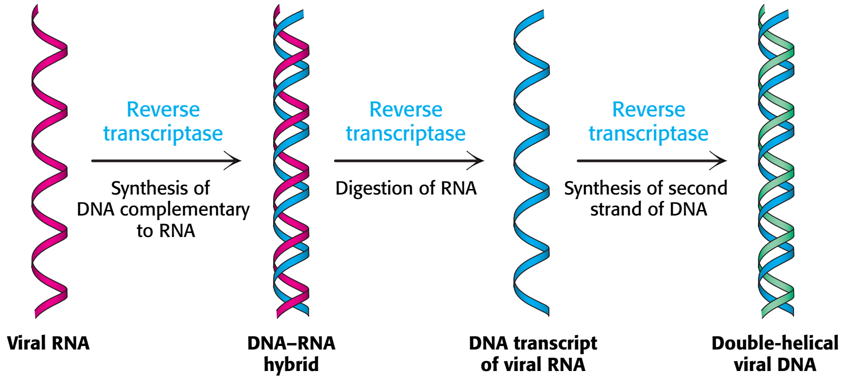

Reverse transcriptase
Synthesis of DNA complementary to RNA
Reverse transcriptase
Digestion of RNA
Reverse transcriptase
Synthesis of second strand of DNA
Viral RNA
DNA–RNA hybrid
DNA transcript of viral RNA
Double-helical viral DNA

FIGURE 4.26 Flow of information from RNA to DNA in retroviruses. The RNA genome of a retrovirus is converted into DNA by reverse transcriptase, an enzyme brought into the cell by the infecting virus particle. Reverse transcriptase possesses several activities and catalyzes the synthesis of a complementary DNA strand, the digestion of the RNA, and the subsequent synthesis of the DNA strand.

susceptible animals. Retrovirus particles contain two copies of a single-stranded RNA molecule. On entering the cell, the RNA is copied into DNA through the action of a viral enzyme called *reverse transcriptase*, which acts as both a polymerase and an RNase (Figure 4.26). The resulting double-helical DNA version of the viral genome can become incorporated into the chromosomal DNA of the host and is replicated along with the normal cellular DNA. At a later time, the integrated viral genome is expressed to form viral RNA and viral proteins, which assemble into new virus particles.

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

The information stored as DNA becomes useful when it is expressed in the production of RNA and proteins. This rich and complex topic is the subject of several chapters later in this book, but here we introduce the basics of gene expression. DNA can be thought of as archival information, stored and manipulated judiciously to minimize damage (mutations). It is expressed in two steps. First, an RNA copy is made that encodes directions for protein synthesis. This messenger RNA can be thought of as a photocopy of the original information: it can be made in multiple copies, used, and then disposed of. Second, the information in messenger RNA is translated to synthesize functional proteins. Other types of RNA molecules exist to facilitate this translation.

Several kinds of RNA play key roles in gene expression

Scientists used to believe that RNA played a passive role in gene expression, as a mere conveyor of information. However, recent investigations have shown that RNA plays a variety of roles, from catalysis to regulation. Cells contain several kinds of RNA (Table 4.3):

TABLE 4.3 RNA molecules in *E. coli*

Type	Relative amount (%)	Sedimentation coefficient (S)	Mass (kDa)	Number of nucleotides
Ribosomal RNA (rRNA)	80	23	1.2×10^3	3700
		16	0.55×10^3	1700
		5	3.6×10^1	120
Transfer RNA (tRNA)	15	4	2.5×10^1	75
Messenger RNA (mRNA)	5		Heterogeneous	

1. *Messenger RNA* (mRNA) is the template for protein synthesis, or *translation*. An mRNA molecule may be produced for each gene or group of genes that is to be expressed in *E. coli,* whereas a distinct mRNA is produced for each gene in eukaryotes. Consequently, mRNA is a heterogeneous class of molecules. In prokaryotes, the average length of an mRNA molecule is about 1.2 kilobases (kb). In eukaryotes, mRNA has structural features, such as stem-loop structures, that regulate the efficiency of translation and the lifetime of the mRNA.

2. *Transfer RNA* (tRNA) carries amino acids in an activated form to the ribosome for peptide-bond formation, in a sequence dictated by the mRNA template. There is at least one kind of tRNA for each of the 20 amino acids. Transfer RNA consists of about 75 nucleotides (having a mass of about 25 kDa).

3. *Ribosomal RNA* (rRNA) is the major component of ribosomes (Chapter 30). In prokaryotes, there are three kinds of rRNA, called *23S*, *16S*, and *5S* RNA because of their sedimentation behavior (p. 74). One molecule of each of these species of rRNA is present in each ribosome. Ribosomal RNA was once believed to play only a structural role in ribosomes. We now know that rRNA is the actual catalyst for protein synthesis.

Kilobase (kb)

A unit of length equal to 1000 base pairs of a double-stranded nucleic acid molecule (or 1000 bases of a single-stranded molecule).

One kilobase of double-stranded DNA has a length of 0.34 μm at its maximal extension (called the contour length) and a mass of about 660 kDa.

Ribosomal RNA is the most abundant of these three types of RNA. Transfer RNA comes next, followed by messenger RNA, which constitutes only 5% of the total RNA. Eukaryotic cells contain additional small RNA molecules that play a variety of roles including the regulation of gene expression, processing of RNA and the synthesis of proteins. We will examine these small RNAs in later chapters. In this chapter, we will consider rRNA, mRNA, and tRNA.

All cellular RNA is synthesized by RNA polymerases

The synthesis of RNA from a DNA template is called *transcription* and is catalyzed by the enzyme *RNA polymerase* (Figure 4.27). RNA polymerase

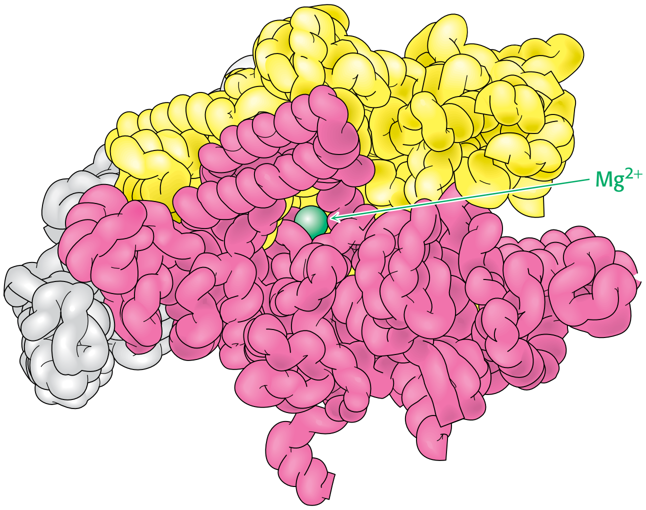


FIGURE 4.27 RNA Polymerase. This large enzyme comprises many subunits, including β (red) and β′ (yellow), which form a “claw” that holds the DNA to be transcribed. *Notice* that the active site includes a Mg^{2+} ion (green) at the center of the structure. The curved tubes making up the protein in the image represent the backbone of the polypeptide chain. [Drawn from 1L9Z, pdb.]

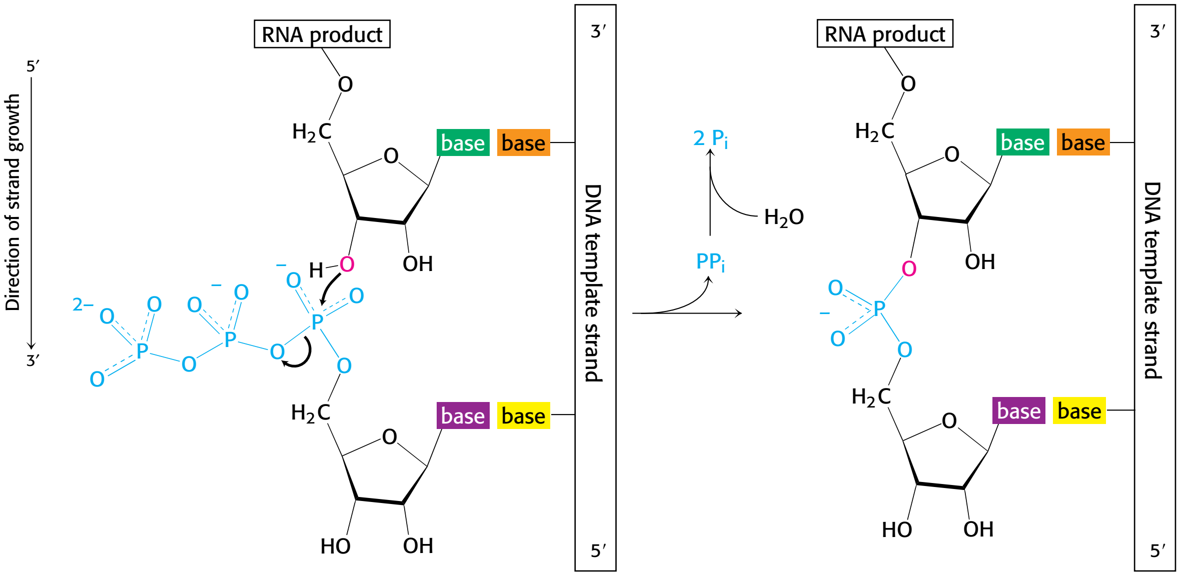


FIGURE 4.28 Transcription mechanism of the chain-elongation reaction catalyzed by RNA polymerase. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 36.3.]

catalyzes the initiation and elongation of RNA chains. The reaction catalyzed by this enzyme is

$$(\text{RNA})_{n\text{ residues}} + \text{ribonucleoside triphosphate} \rightleftharpoons (\text{RNA})_{n+1\text{ residues}} + \text{PP}_i$$

RNA polymerase requires the following components:

1. *A template.* The preferred template is *double-stranded DNA.* Single-stranded DNA also can serve as a template. RNA, whether single or double stranded, is not an effective template; nor are RNA–DNA hybrids.
2. *Activated precursors.* All four *ribonucleoside triphosphates*—ATP, GTP, UTP, and CTP—are required.
3. *A divalent metal ion.* Either Mg^{2+} or Mn^{2+} is effective.

The synthesis of RNA is like that of DNA in several respects (Figure 4.28). First, the direction of synthesis is 5′ → 3′. Second, the mechanism of elongation is similar: the 3′-OH group at the terminus of the growing chain makes a nucleophilic attack on the innermost phosphoryl group of the incoming nucleoside triphosphate. Third, the synthesis is driven forward by the hydrolysis of pyrophosphate. In contrast with DNA polymerase, however, RNA polymerase does not require a primer. In addition, the ability of RNA polymerase to correct mistakes is not as extensive as that of DNA polymerase.

All three types of cellular RNA—mRNA, tRNA, and rRNA—are synthesized in *E. coli* by the same RNA polymerase according to instructions given by a DNA template. In mammalian cells, there is a division of labor among several different kinds of RNA polymerases. We shall return to these RNA polymerases in Chapter 29.

Table 4.4 Base composition (percentage) of RNA synthesized from a viral DNA template

DNA template (plus, or coding, strand of ϕX174)		RNA product	
A	25	U	25
T	33	A	32
G	24	C	23
C	18	G	20

RNA polymerases take instructions from DNA templates

RNA polymerase, like the DNA polymerases described earlier, takes instructions from a DNA template. The earliest evidence was the finding that the *base composition* of newly synthesized RNA is the complement of that of the DNA template strand, as exemplified by the RNA synthesized from a template of single-stranded DNA from the ϕX174 virus (Table 4.4). The strongest evidence for the fidelity of transcription came from base-sequence studies. For instance, the nucleotide sequence of a segment of the gene encoding the enzymes required for tryptophan synthesis was determined with the use of DNA-sequencing techniques (Section 5.1). Likewise, the sequence of the mRNA for the corresponding gene was determined. The results showed that the RNA sequence is the precise complement of the DNA template sequence (Figure 4.29).

5′—GCGGCGACGCGCAGUUAAUCCCACAGCCGCCAGUUCCGCUGGCGGCAU—3′ **mRNA**
3′—CGCCGCTGCGCGTCAATTAGGGTGTCGGCGGTCAAGGCGACCGCCGTA—5′ **Template strand of DNA**
5′—GCGGCGACGCGCAGTTAATCCCACAGCCGCCAGTTCCGCTGGCGGCAT—3′ **Coding strand of DNA**

FIGURE 4.29 **Complementarity between mRNA and DNA.** The base sequence of mRNA (red) is the complement of that of the DNA template strand (blue). The sequence shown here is from the tryptophan operon, a segment of DNA containing the genes for five enzymes that catalyze the synthesis of tryptophan. The other strand of DNA (black) is called the coding strand because it has the same sequence as the RNA transcript except for thymine (T) in place of uracil (U).

Consensus sequence

Not all base sequences of promoter sites are identical. However, they do possess common features, which can be represented by an idealized consensus sequence. Each base in the consensus sequence TATAAT is found in most prokaryotic promoters. Nearly all promoter sequences differ from this consensus sequence at only one or two bases.

Transcription begins near promoter sites and ends at terminator sites

RNA polymerase must detect and transcribe discrete genes from within large stretches of DNA. What marks the beginning of the unit to be transcribed? DNA templates contain regions called *promoter sites* that specifically bind RNA polymerase and determine where transcription begins. In bacteria, two sequences on the 5′ (upstream) side of the first nucleotide to be transcribed function as promoter sites (Figure 4.30A). One of them, called the *Pribnow box*, has the consensus sequence TATAAT and is centered at −10 (10 nucleotides on the 5′ side of the first nucleotide transcribed, which is denoted by +1). The other, called the −35 *region*, has the consensus sequence TTGACA. The first nucleotide transcribed is usually a purine.

Eukaryotic genes encoding proteins have promoter sites with a TATAAA consensus sequence, called a *TATA box* or a *Hogness box*, centered at about −25 (Figure 4.30B). Many eukaryotic promoters also have a *CAAT box* with a GGN*CAAT*CT consensus sequence centered at about −75. The transcription of eukaryotic genes is further stimulated by *enhancer sequences*, which can be quite distant (as many as several kilobases) from the start site, on either its 5′ or its 3′ side.

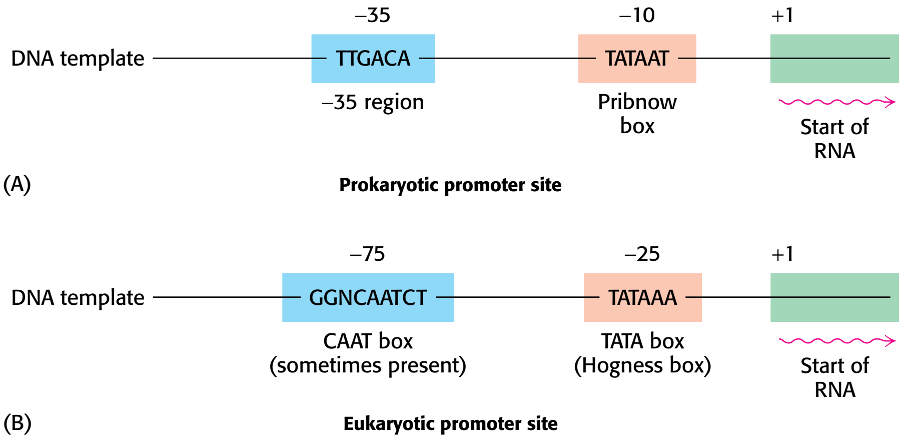


FIGURE 4.30 **Promoter sites for transcription in (A) prokaryotes and (B) eukaryotes.** Consensus sequences are shown. The first nucleotide to be transcribed is numbered +1. The adjacent nucleotide on the 5′ side is numbered −1. The sequences shown are those of the coding strand of DNA.

In *E. coli*, RNA polymerase proceeds along the DNA template, transcribing one of its strands until it synthesizes a terminator sequence. This sequence encodes a termination signal, which is a *base-paired hairpin* on the newly synthesized RNA molecule (Figure 4.31). This hairpin is formed by base-pairing of self-complementary sequences that are rich in G and C. Nascent RNA spontaneously dissociates from RNA polymerase when this hairpin is followed by a string of U residues. Alternatively, RNA synthesis can be terminated by the action of *rho,* a protein. Less is known about the termination of transcription in eukaryotes. A more detailed discussion of the initiation and termination of transcription will be given in Chapter 29. The important point now is that *discrete start and stop signals for transcription are encoded in the DNA template.*

In eukaryotes, the messenger RNA is modified after transcription (Figure 4.32). A "cap" structure, a guanosine nucleotide attached to the mRNA with an unusual 5′-5′ triphosphate linkage, is attached to the 5′ end, and a sequence of adenylates, the poly(A) tail, is added to the 3′ end. These modifications will be presented in detail in Chapter 29.

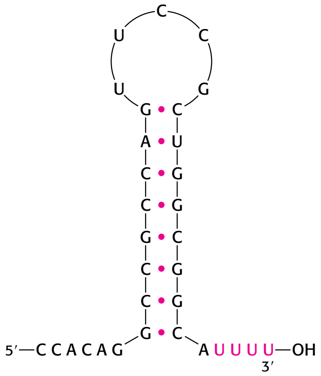


FIGURE 4.31 Base sequence of the 3′ end of an mRNA transcript in *E. coli.* A stable hairpin structure is followed by a sequence of uridine (U) residues.

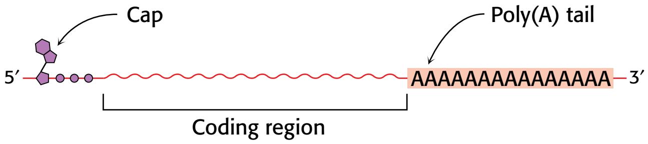


FIGURE 4.32 Modification of mRNA. Messenger RNA in eukaryotes is modified after transcription. A nucleotide "cap" structure is added to the 5′ end, and a poly(A) tail is added at the 3′ end.

Transfer RNAs are the adaptor molecules in protein synthesis

We have seen that mRNA is the template for protein synthesis. How then does it direct amino acids to become joined in the correct sequence to form a protein? In 1958, Francis Crick wrote:

> RNA presents mainly a sequence of sites where hydrogen bonding could occur. One would expect, therefore, that whatever went onto the template in a *specific* way did so by forming hydrogen bonds. It is therefore a natural hypothesis that the amino acid is carried to the template by an adaptor molecule, and that the adaptor is the part that actually fits onto the RNA. In its simplest form, one would require twenty adaptors, one for each amino acid.

This highly innovative hypothesis soon became established as fact. *The adaptors in protein synthesis are transfer RNAs*. The structure and reactions of these remarkable molecules will be considered in detail in Chapter 30. For the moment, it suffices to note that tRNAs contain an *amino acid-attachment site* and a *template-recognition site.* A tRNA molecule carries a specific amino acid in an activated form to the ribosome. The carboxyl group of this amino acid is esterified to the 3′- or 2′-hydroxyl group of the ribose unit of an adenylate at the 3′ end of the tRNA molecule. The adenylate is always preceded by two cytidylates to form the CCA arm of the tRNA (Figure 4.33). The joining of an amino acid to a tRNA molecule to form an *aminoacyl-tRNA* is catalyzed by a specific enzyme called an *aminoacyl-tRNA synthetase*. This esterification reaction is driven by ATP cleavage. There is at least one specific synthetase for each of the 20 amino acids. The template-recognition site on tRNA is a sequence of three bases called an *anticodon* (Figure 4.34). The anticodon on tRNA recognizes a complementary sequence of three bases, called a *codon*, on mRNA.

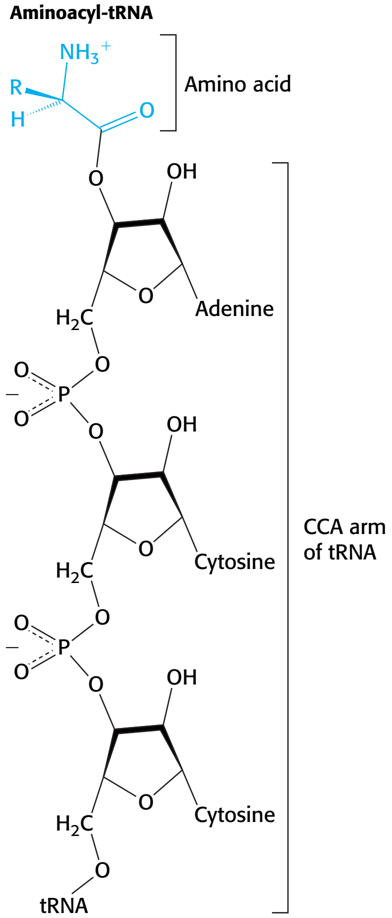


FIGURE 4.33 Attachment of an amino acid to a tRNA molecule. The amino acid (shown in blue) is esterified to the 3′-hydroxyl group of the terminal adenylate of tRNA. [Source: J. L. Tymoczko, J. Berg, and L. Stryer, *Biochemistry: A Short Course*, 2nd ed. (W. H. Freeman and Company, 2013), Fig. 39.3.]

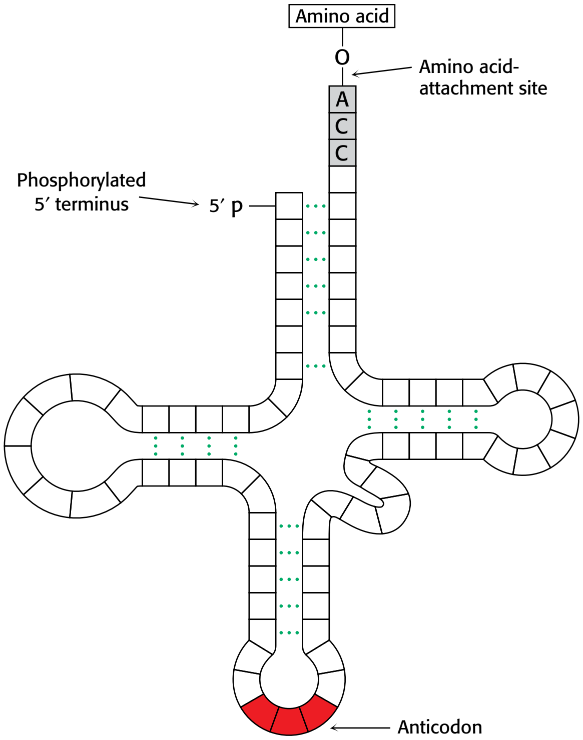


FIGURE 4.34 General structure of an aminoacyl-tRNA. The amino acid is attached at the 3′ end of the RNA. The anticodon is the template-recognition site. *Notice* that the tRNA has a cloverleaf structure with many hydrogen bonds (green dots) between bases.

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

The *genetic code* is the relation between the sequence of bases in DNA (or its RNA transcripts) and the sequence of amino acids in proteins. Experiments by Marshall Nirenberg, Har Gobind Khorana, Francis Crick, Sydney Brenner, and others established the following features of the genetic code by 1961:

1. *Three nucleotides encode an amino acid*. Proteins are built from a basic set of 20 amino acids, but there are only four bases. Simple calculations show that a minimum of three bases is required to encode at least 20 amino acids. Genetic experiments showed that *an amino acid is in fact encoded by a group of three bases, or codon*.

2. *The code is nonoverlapping*. Consider a base sequence ABCDEF. In an overlapping code, ABC specifies the first amino acid, BCD the next, CDE the next, and so on. In a nonoverlapping code, ABC designates the first amino acid, DEF the second, and so forth. Genetic experiments again established the code to be nonoverlapping.

3. *The code has no punctuation*. In principle, one base (denoted as Q) might serve as a "comma" between groups of three bases.

. . . QABCQDEFQGHIQJKLQ . . .

However, it is not the case. Rather, *the sequence of bases is read sequentially from a fixed starting point*, without punctuation.

4. *The code has directionality*. The code is read from the 5′ end of the messenger RNA to its 3′ end.

5. *The genetic code is degenerate*. Most amino acids are encoded by more than one codon. There are 64 possible base triplets and only 20 amino acids, and in fact 61 of the 64 possible triplets specify particular amino acids. Three triplets (called *stop codons*) designate the termination of translation. Thus, *for most amino acids, there is more than one code word.*

Major features of the genetic code

All 64 codons have been deciphered (Table 4.5). Because the code is highly degenerate, only tryptophan and methionine are encoded by just one triplet each. Each of the other 18 amino acids is encoded by two or more. Indeed, leucine, arginine, and serine are specified by six codons each.

Codons that specify the same amino acid are called *synonyms*. For example, CAU and CAC are synonyms for histidine. Note that synonyms are not distributed haphazardly throughout the genetic code. In Table 4.5, an amino acid specified by two or more synonyms occupies a single box (unless it is specified by more than four synonyms). The amino acids in a box are specified by codons that have the same first two bases but differ in the third base, as exemplified by GUU, GUC, GUA, and GUG. Thus, *most synonyms differ only in the last base of the triplet*. Inspection of the code shows that XYC and XYU always encode the same amino acid, and XYG and XYA usually encode the same amino acid as well. The structural basis for these equivalences of codons becomes evident when we consider the nature of the anticodons of tRNA molecules (Section 30.3).

What is the biological significance of the extensive degeneracy of the genetic code? If the code were not degenerate, 20 codons would designate amino acids and 44 would lead to chain termination. The probability of mutating to chain termination would therefore be much higher with a nondegenerate code. Chain-termination mutations usually lead to inactive proteins, whereas substitutions of one amino acid for another are usually rather harmless. Moreover, the code is constructed such that a change in any single nucleotide base of a codon results in a synonym or an amino acid with similar chemical properties. Thus, *degeneracy minimizes the deleterious effects of mutations*.

TABLE 4.5 The genetic code

First Position (5′ end)	Second Position U	C	A	G	Third Position (3′ end)
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr Stop Stop	Cys Cys Stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Note: This table identifies the amino acid encoded by each triplet. For example, the codon 5′- AUG-3′ on mRNA specifies methionine, whereas CAU specifies histidine. UAA, UAG, and UGA are termination signals. AUG is part of the initiation signal, in addition to coding for internal methionine residues.

Messenger RNA contains start and stop signals for protein synthesis

Messenger RNA is translated into proteins on *ribosomes*—large molecular complexes assembled from proteins and ribosomal RNA. How is mRNA interpreted by the translation apparatus? The start signal for protein synthesis is complex in bacteria. Polypeptide chains in bacteria start with a modified amino acid—namely, formylmethionine (fMet). A specific tRNA, the initiator tRNA, carries fMet. This fMet-tRNA recognizes the codon AUG. However, AUG is also the codon for an internal methionine residue. Hence, the signal for the first amino acid in a prokaryotic polypeptide chain must be more complex than that for all subsequent ones. *AUG is only part of the initiation signal* (Figure 4.35). In bacteria, the initiating AUG codon is preceded several nucleotides away by a purine-rich sequence, called the *Shine–Dalgarno sequence,* that base-pairs with a complementary sequence in a ribosomal RNA molecule (Section 30.3). In eukaryotes, the AUG closest to the 5′ end of an mRNA molecule is usually the start signal for protein synthesis. This particular AUG is read by an initiator tRNA conjugated to methionine. After the initiator AUG has been located, the *reading frame* is established—groups of three nonoverlapping nucleotides are defined, beginning with the initiator AUG codon.

As already mentioned, *UAA, UAG, and UGA designate chain termination*. These codons are read not by tRNA molecules but rather by specific proteins called *release factors* (Section 30.3). Binding of a release factor to the ribosome releases the newly synthesized protein.

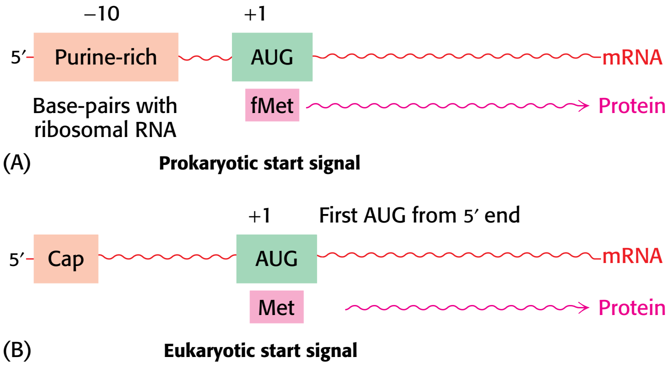


FIGURE 4.35 Initiation of protein synthesis. Start signals are required for the initiation of protein synthesis in (A) prokaryotes and (B) eukaryotes.

The genetic code is nearly universal

Most organisms use the same genetic code. This universality accounts for the fact that human proteins, such as insulin, can be synthesized in the bacterium *E. coli* and harvested from it for the treatment of diabetes. However, genome-sequencing studies have shown that not all genomes are translated by the same code. Ciliated protozoa, for example, differ from most organisms in that UAA and UAG are read as codons for amino acids rather than as stop signals; UGA is their sole termination signal. The first variations in the genetic code were found in mitochondria from a number of species, including human beings (Table 4.6). The genetic code of mitochondria can differ from that of the rest of the cell because mitochondrial DNA encodes a distinct set of transfer RNAs, adaptor molecules that recognize the alternative codons. Thus, the genetic code is nearly but not absolutely universal.

Why has the code remained nearly invariant through billions of years of evolution, from bacteria to human beings? A mutation that altered the reading of mRNA would change the amino acid sequence of most, if not all,

TABLE 4.6 Distinctive codons of human mitochondria

Codon	Standard code	Mitochondrial code
UGA	Stop	Trp
UGG	Trp	Trp
AUA	Ile	Met
AUG	Met	Met
AGA	Arg	Stop
AGG	Arg	Stop

proteins synthesized by that particular organism. Many of these changes would undoubtedly be deleterious, and so there would be strong selection against a mutation with such pervasive consequences.

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons

In bacteria, polypeptide chains are encoded by a continuous array of triplet codons in DNA. For many years, genes in higher organisms were assumed to be organized in the same manner. This view was unexpectedly shattered in 1977, when investigators discovered that most eukaryotic genes are *discontinuous*. The mosaic nature of eukaryotic genes was revealed by electron microscopic studies of hybrids formed between mRNA and a segment of DNA containing the corresponding gene (Figure 4.36). For example, the gene for the β chain of hemoglobin is interrupted within its amino acid-coding sequence by a long stretch of of 550 non-coding base pairs and a short one of 120 base pairs. Thus, the *β-globin gene is split into three coding sequences* (Figure 4.37). Non-coding regions are called *introns* (for *int*ervening sequences), whereas coding regions are called *exons* (for *ex*pressed sequences). The average human gene has 8 introns, and some have more than 100. The size ranges from 50 to 10,000 nucleotides.

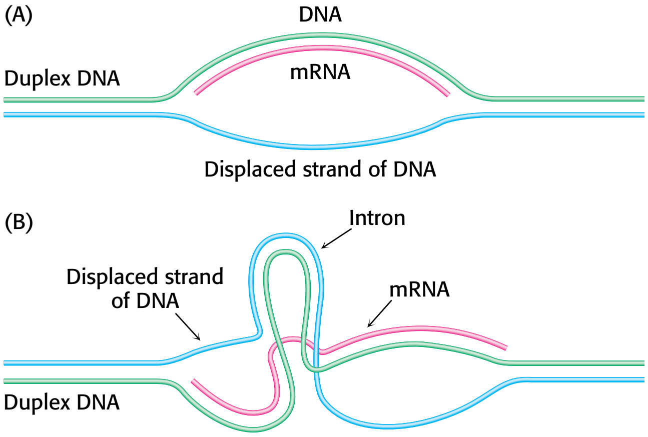


FIGURE 4.36 Detection of introns by electron microscopy. An mRNA molecule (shown in red) is hybridized to genomic DNA containing the corresponding gene. (A) A single loop of single-stranded DNA (shown in blue) is seen if the gene is continuous. (B) Two loops of single-stranded DNA (blue) and a loop of double-stranded DNA (blue and green) are seen if the gene contains an intron. Additional loops are evident if more than one intron is present.

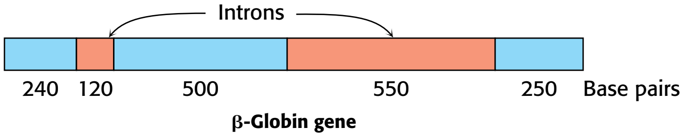


FIGURE 4.37 Structure of the β-globin gene.

RNA processing generates mature RNA

At what stage in gene expression are introns removed? Newly synthesized RNA molecules (pre-mRNA or primary transcript) isolated from nuclei are much larger than the mRNA molecules derived from them; in regard to β-globin RNA, the former consists of approximately 1600 nucleotides and the latter approximately 900 nucleotides. In fact, the primary transcript of the β-globin gene contains two regions that are not present in the mRNA. *These regions in primary transcript are excised, and the coding sequences are simultaneously linked by a precise splicing complex to form the mature mRNA*

(Figure 4.38). A common feature in the expression of discontinuous, or split, genes is that their exons are ordered in the same sequence in mRNA as in DNA. Thus, the codons in split genes, like continuous genes, are in the same linear order as the amino acids in the polypeptide products.

Splicing is a complex operation that is carried out by *spliceosomes,* which are assemblies of proteins and small RNA molecules (snRNA). RNA plays the catalytic role (Section 29.3). Spliceosomes recognize signals in the nascent RNA that specify the splice sites. *Introns nearly always begin with GU and end with an AG that is preceded by a pyrimidine-rich tract* (Figure 4.39). *This consensus sequence is part of the signal for splicing.*

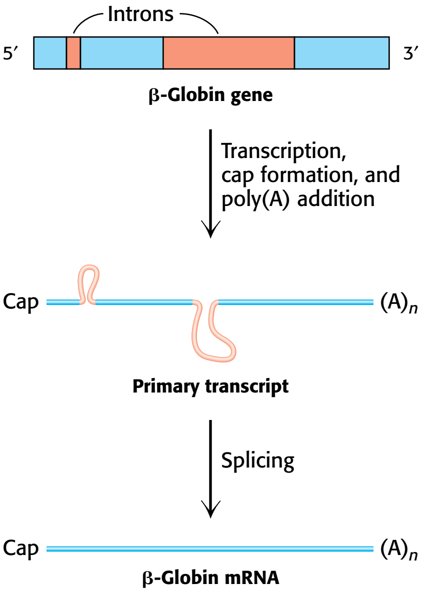


FIGURE 4.38 Transcription and processing of the β-globin gene. The gene is transcribed to yield the primary transcript, which is modified by cap and poly(A) addition. The introns in the primary RNA transcript are removed to form the mRNA.

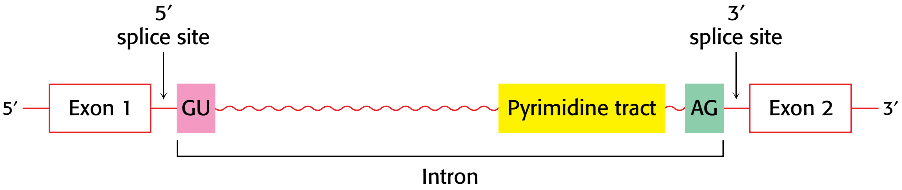


FIGURE 4.39 Consensus sequence for the splicing of mRNA precursors.

Many exons encode protein domains

Most genes of higher eukaryotes, such as birds and mammals, are split. Lower eukaryotes, such as yeast, have a much higher proportion of continuous genes. In prokaryotes, split genes are extremely rare. Have introns been inserted into genes in the evolution of higher organisms? Or have introns been removed from genes to form the streamlined genomes of prokaryotes and simple eukaryotes? Comparisons of the DNA sequences of genes encoding evolutionarily conserved proteins suggest that *introns were present in ancestral genes and were lost in the evolution of organisms that have become optimized for very rapid growth, such as prokaryotes.* The positions of introns in some genes are at least 1 billion years old. Furthermore, a common mechanism of splicing developed before the divergence of fungi, plants, and vertebrates, as shown by the finding that mammalian cell extracts can splice yeast RNA.

What advantages might split genes confer? *Many exons encode discrete structural and functional domains of proteins.* An attractive hypothesis is that new *proteins arose in evolution by the rearrangement of exons encoding discrete structural elements, binding sites, and catalytic sites,* a process called *exon shuffling.* Because it preserves functional units but allows them to interact in new ways, exon shuffling is a rapid and efficient means of generating novel genes (Figure 4.40). Figure 4.41 shows the composition of a gene that was formed in part by exon shuffling. DNA can break and recombine in introns with no deleterious effect on encoded proteins. In contrast, the exchange of sequences within different exons usually leads to loss of function.

Another advantage of split genes is the potential for generating a series of related proteins by alternative splicing of the primary transcript. For example, a precursor of an antibody-producing cell forms an antibody that is anchored in the cell's plasma membrane (Figure 4.42). The attached antibody recognizes a specific foreign antigen, an event that leads to cell differentiation and proliferation. The activated antibody-producing cells then splice their primary transcript in an alternative manner to form soluble antibody molecules that are secreted rather than retained on the cell surface. *Alternative splicing is a facile means of forming a set of proteins that are variations of a basic motif without requiring a gene for each protein.* Because of alternative splicing, the proteome is more diverse than the genome in eukaryotes.

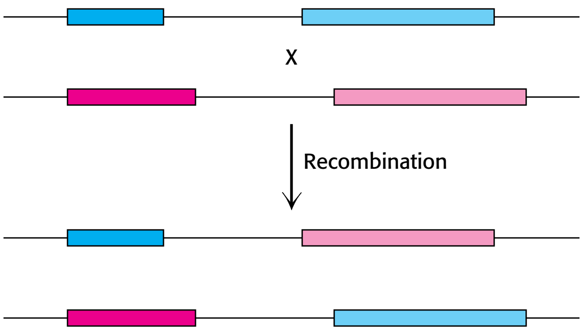


FIGURE 4.40 Exon shuffling. Exons can be readily shuffled by recombination of DNA to expand the genetic repertoire.

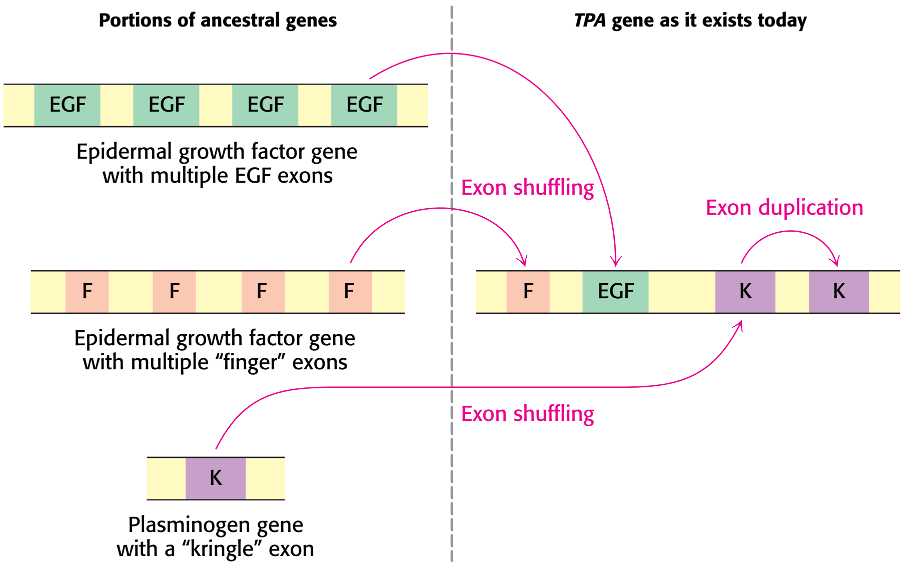

Portions of ancestral genes
TPA gene as it exists today
EGF
EGF
EGF
EGF
Epidermal growth factor gene with multiple EGF exons
Exon shuffling
Exon duplication
F
F
F
F
F
EGF
K
K
Epidermal growth factor gene with multiple "finger" exons
Exon shuffling
K
Plasminogen gene with a "kringle" exon

FIGURE 4.41 The tissue plasminogen activator (TPA) gene was generated by exon shuffling. The gene for TPA encodes an enzyme that functions in hemostasis (Section 10.4). This gene consists of 4 exons, one (F) derived from the fibronectin gene which encodes an extracellular matrix protein, one from the epidermal growth factor gene (EGF), and two from the plasminogen gene (K, Section 10.4), the substrate of the TPA protein. The K domain appears to have arrived by exon shuffling and then been duplicated to generate the TPA gene that exists today. [Information from: www.ehu.es/ehusfera/genetica/2012/10/02/demostracion-molecular-de-microevolucion/]

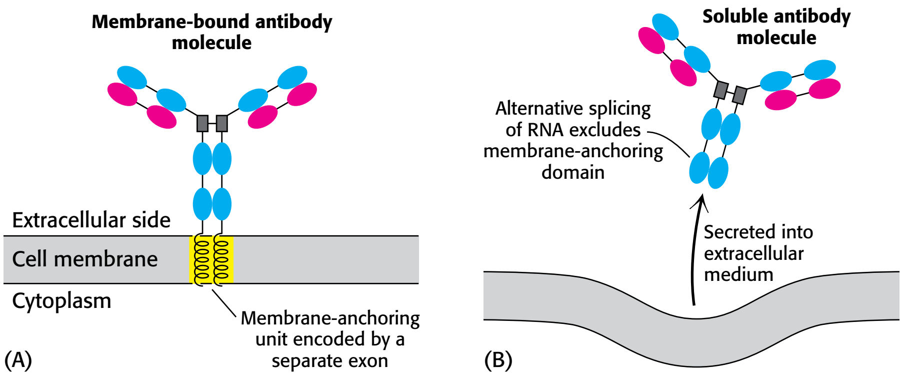

Membrane-bound antibody molecule
Extracellular side
Cell membrane
Cytoplasm
Membrane-anchoring unit encoded by a separate exon
(A)
Soluble antibody molecule
Alternative splicing of RNA excludes membrane-anchoring domain
Secreted into extracellular medium
(B)

FIGURE 4.42 Alternative splicing. Alternative splicing generates mRNAs that are templates for different forms of a protein: (A) a membrane-bound antibody on the surface of a lymphocyte and (B) its soluble counterpart, exported from the cell. The membrane-bound antibody is anchored to the plasma membrane by a helical segment (highlighted in yellow) that is encoded by its own exon.

SUMMARY

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

DNA and RNA are linear polymers of a limited number of monomers. In DNA, the repeating units are nucleotides, with the sugar being a deoxyribose and the bases being adenine (A), thymine (T), guanine (G), and cytosine (C). In RNA, the sugar is a ribose and the base uracil (U) is used in place of thymine. DNA is the molecule of heredity in all prokaryotic and eukaryotic organisms. In viruses, the genetic material is either DNA or RNA.

4.2 A Pair of Nucleic Acid Strands with Complementary Sequences Can Form a Double-Helical Structure

All cellular DNA consists of two very long, helical polynucleotide strands coiled around a common axis. The sugar–phosphate backbone of each strand is on the outside of the double helix, whereas the purine and pyrimidine bases are on the inside stabilized by stacking forces. The two strands are held together by hydrogen bonds between pairs of bases: adenine is always paired with thymine, and guanine is always paired with cytosine. Hence, one strand of a double helix is the complement of the other. The two strands of the double helix run in opposite directions. Genetic information is encoded in the precise sequence of bases along a strand.

DNA is a structurally dynamic molecule that can exist in a variety of helical forms: A-DNA, B-DNA (the classic Watson–Crick helix), and Z-DNA. In A-, B-, and Z-DNA, two antiparallel chains are held together by Watson–Crick base pairs and stacking interactions between bases in the same strand. A- and B-DNA are right-handed helices. In B-DNA, the base pairs are nearly perpendicular to the helix axis. Z-DNA is a left-handed helix. Most of the DNA in a cell is in the B-form.

Double-stranded DNA can also wrap around itself to form a supercoiled structure. The supercoiling of DNA has two important consequences. Supercoiling compacts the DNA and, because supercoiled DNA is partly unwound, it is more accessible for interactions with other biomolecules.

Single-stranded nucleic acids, most notably RNA, can form complicated three-dimensional structures that may contain extensive double-helical regions that arise from the folding of the chain into hairpins.

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

The structural nature of the double helix readily accounts for the accurate replication of genetic material because the sequence of bases in one strand determines the sequence of bases in the other strand. In replication, the strands of the helix separate and a new strand complementary to each of the original strands is synthesized. Thus, two new double helices are generated, each composed of one strand from the original molecule and one newly synthesized strand. This mode of replication is called semiconservative replication because each new helix retains one of the original strands.

In order for replication to take place, the strands of the double helix must be separated. In vitro, heating a solution of double-helical DNA separates the strands, a process called melting. On cooling, the strands reanneal and re-form the double helix. In the cell, special proteins temporarily separate the strands in replication.

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

In the replication of DNA, the two strands of a double helix unwind and separate as new strands are synthesized. Each parent strand, with the help of a primer, acts as a template for the formation of a new complementary strand. The replication of DNA is a complex process carried out by many proteins, including several DNA polymerases. The activated precursors in the synthesis of DNA are the four deoxyribonucleoside 5′-triphosphates. The new strand is synthesized in the 5′ → 3′ direction by a nucleophilic attack by the 3′-hydroxyl terminus of the primer strand on the innermost phosphorus atom of the incoming deoxyribonucleoside triphosphate. Most important, DNA polymerases catalyze the formation of a phosphodiester linkage only if the base on the incoming nucleotide is complementary to the base on the template strand. In other words, DNA polymerases are template-directed enzymes. The genes of some viruses, such as tobacco mosaic virus, are made of single-stranded RNA. An RNA-directed RNA polymerase mediates the replication of this viral RNA. Retroviruses, exemplified by HIV-1, have a single-stranded RNA genome that undergoes reverse transcription into double-stranded DNA by reverse transcriptase, an RNA-directed DNA polymerase.

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

The flow of genetic information in normal cells is from DNA to RNA to protein. The synthesis of RNA from a DNA template is called transcription, whereas the synthesis of a protein from an RNA template is termed translation. Cells contain several kinds of RNA, among which are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), which vary in size from 75 to more than 5000 nucleotides. All cellular RNA is synthesized by RNA polymerases according to instructions given by DNA templates. The activated intermediates are ribonucleoside triphosphates and the direction of synthesis, like that of DNA, is $5' \rightarrow 3'$. RNA polymerase differs from DNA polymerase in not requiring a primer.

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

The genetic code is the relation between the sequence of bases in DNA (or its RNA transcript) and the sequence of amino acids in proteins. Amino acids are encoded by groups of three bases (called codons) starting from a fixed point. Sixty-one of the 64 codons specify particular amino acids, whereas the other 3 codons (UAA, UAG, and UGA) are signals for chain termination. Thus, for most amino acids, there is more than one code word. In other words, the code is degenerate. The genetic code is nearly the same in all organisms. Natural mRNAs contain start and stop signals for translation, just as genes do for directing where transcription begins and ends.

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons

Most genes in higher eukaryotes are discontinuous. Coding sequences in these split genes, called exons, are separated by noncoding sequences, called introns, which are removed in the conversion of the primary transcript into mRNA and other functional mature RNA molecules. Split genes, like continuous genes, are colinear with their polypeptide products. A striking feature of many exons is that they encode functional domains in proteins. New proteins probably arose in the course of evolution by the shuffling of exons. Introns may have been present in primordial genes but were lost in the evolution of such fast-growing organisms as bacteria and yeast.

KEY TERMS

double helix (p. 105)
deoxyribonucleic acid (DNA) (p. 106)
deoxyribose (p. 106)
ribose (p. 106)
purine (p. 107)
pyrimidine (p. 107)
ribonucleic acid (RNA) (p. 107)
nucleoside (p. 107)
nucleotide (p. 107)
B-DNA (p. 111)
A-DNA (p. 111)
Z-DNA (p. 112)
semiconservative replication (p. 114)
DNA polymerase (p. 117)
template (p. 117)
primer (p. 118)
reverse transcriptase (p. 119)
messenger RNA (mRNA) (p. 120)
translation (p. 120)
transfer RNA (tRNA) (p. 120)
ribosomal RNA (rRNA) (p. 120)
transcription (p. 120)
RNA polymerase (p. 120)
promoter site (p. 122)
anticodon (p. 123)
codon (p. 123)
genetic code (p. 124)
ribosome (p. 126)
Shine–Dalgarno sequence (p. 126)
intron (p. 127)
exon (p. 127)
splicing (p. 128)
spliceosomes (p. 128)
exon shuffling (p. 128)
alternative splicing (p. 128)

PROBLEMS

1. *A t instead of an s?* Differentiate between a nucleoside and a nucleotide.

2. *A lovely pair*. What is a Watson–Crick base pair?

3. *Chargaff rules!* Biochemist Erwin Chargaff was the first to note that, in DNA, [A] = [T] and [G] = [C], equalities now called Chargaff's rule. Using this rule, determine the percentages of all the bases in DNA that is 20% thymine.

4. *But not always*. A single strand of RNA is 20% U. What can you predict about the percentages of the remaining bases?

5. *Complements.* Write the complementary sequence (in the standard $5' \rightarrow 3'$ notation) for (a) GATCAA, (b) TCGAAC, (c) ACGCGT, and (d) TACCAT.

6. *Compositional constraint.* The composition (in mole-fraction units) of one of the strands of a double-helical DNA molecule is $[A] = 0.30$ and $[G] = 0.24$. (a) What can you say about [T] and [C] for the same strand? (b) What can you say about [A], [G], [T], and [C] of the complementary strand?

7. *Size matters*. Why are GC and AT the only base pairs permissible in the double helix?

8. *Strong, but not strong enough*. Why does heat denature, or melt, DNA in solution?

9. *Uniqueness.* The human genome contains 3 billion nucleotides arranged in a vast array of sequences. What is the minimum length of a DNA sequence that will, in all probability, appear only once in the human genome? You need consider only one strand and may assume that all four nucleotides have the same probability of appearance.

10. *Coming and going*. What does it mean to say that the DNA strands in a double helix have opposite directionality or polarity?

11. *All for one*. If the forces—hydrogen bonds and stacking forces—holding a helix together are weak, why is it difficult to disrupt a double helix?

12. *Overcharged*. DNA in the form of a double helix must be associated with cations, usually Mg^{2+}. Why is this requirement the case?

13. *Not quite from A to Z*. Describe the three forms that a double helix can assume.

14. *Lost DNA*. The DNA of a deletion mutant of λ bacteriophage has a length of 15 μm instead of 17 μm. How many base pairs are missing from this mutant?

15. *Axial ratio*. What is the axial ratio (length:diameter) of a DNA molecule 20 μm long?

16. *Guide and starting point*. Define template and primer as they relate to DNA synthesis.

17. *An unseen pattern.* What result would Meselson and Stahl have obtained if the replication of DNA were conservative (i.e., the parental double helix stayed together)? Give the expected distribution of DNA molecules after 1.0 and 2.0 generations for conservative replication.

18. *Which way?* Explain, on the basis of nucleotide structure, why DNA synthesis proceeds in the 5′-to-3′ direction.

19. *Tagging DNA*.

(a) Suppose that you want to radioactively label DNA but not RNA in dividing and growing bacterial cells. Which radioactive molecule would you add to the culture medium?
(b) Suppose that you want to prepare DNA in which the backbone phosphorus atoms are uniformly labeled with ^{32}P. Which precursors should be added to a solution containing DNA polymerase and primed template DNA? Specify the position of radioactive atoms in these precursors.

20. *Finding a template.* A solution contains DNA polymerase and the Mg^{2+} salts of dATP, dGTP, dCTP, and TTP. The following DNA molecules are added to aliquots of this solution. Which of them would lead to DNA synthesis?

(a) A single-stranded closed circle containing 1000 nucleotide units.
(b) A double-stranded closed circle containing 1000 nucleotide pairs.
(c) A single-stranded closed circle of 1000 nucleotides base-paired to a linear strand of 500 nucleotides with a free 3′-OH terminus.
(d) A double-stranded linear molecule of 1000 nucleotide pairs with a free 3′-OH group at each end.

21. *Retrograde*. What is a retrovirus and how does information flow for a retrovirus differ from that for the infected cell?

22. *The right start.* Suppose that you want to assay reverse transcriptase activity. If polyriboadenylate is the template in the assay, what should you use as the primer? Which radioactive nucleotide should you use to follow chain elongation?

23. *Essential degradation.* Reverse transcriptase has ribonuclease activity as well as polymerase activity. What is the role of its ribonuclease activity?

24. *Virus hunting.* You have purified a virus that infects turnip leaves. Treatment of a sample with phenol removes viral proteins. Application of the residual material to scraped leaves results in the formation of progeny virus particles. You infer that the infectious substance is a nucleic acid. Propose a simple and highly sensitive means of determining whether the infectious nucleic acid is DNA or RNA.

25. *Mutagenic consequences.* Spontaneous deamination of cytosine bases in DNA takes place at low but measurable frequency. Cytosine is converted into uracil by loss of its amino group. After this conversion, which base pair occupies this position in each of the daughter strands resulting from one round of replication? Two rounds of replication?

26. *Information content.*
(a) How many different 8-mer sequences of DNA are there? (Hint: There are 16 possible dinucleotides and 64 possible trinucleotides.) We can quantify the information-carrying capacity of nucleic acids in the following way. Each position can be one of four bases, corresponding to two bits of

information ($2^2 = 4$). Thus, a chain of 5100 nucleotides corresponds to $2 \times 5100 = 10{,}200$ bits, or 1275 bytes (1 byte = 8 bits).
(b) How many bits of information are stored in an 8-mer DNA sequence? In the *E. coli* genome? In the human genome?
(c) Compare each of these values with the amount of information that can be stored on a computer compact disc, or CD (about 700 megabytes).

27. *Key polymerases.* Compare DNA polymerase and RNA polymerase from *E. coli* in regard to each of the following features: (a) activated precursors, (b) direction of chain elongation, (c) conservation of the template, and (d) need for a primer.

28. *Different strands.* Explain the difference between the coding strand and the template strand in DNA.

29. *Family resemblance.* Differentiate among mRNA, rRNA and tRNA.

30. *A code you can live by.* What are the key characteristics of the genetic code?

31. *Encoded sequences.*

(a) Write the sequence of the mRNA molecule synthesized from a DNA template strand having the following sequence:

5′–ATCGTACCGTTA–3′

(b) What amino acid sequence is encoded by the following base sequence of an mRNA molecule? Assume that the reading frame starts at the 5′ end.

5′–UUGCCUAGUGAUUGGAUG–3′

(c) What is the sequence of the polypeptide formed on addition of poly(UUAC) to a cell-free protein-synthesizing system?

32. *A tougher chain.* RNA is readily hydrolyzed by alkali, whereas DNA is not. Why?

33. *A picture is worth a thousand words.* Write a reaction sequence showing why RNA is more susceptible to nucleophilic attack than DNA.

34. *Flowing information.* What is meant by the phrase gene expression?

35. *We can all agree on that.* What is a consensus sequence?

36. *A potent blocker.* Cordycepin (3′-deoxyadenosine) is an adenosine analog. When converted into cordycepin 5′-triphosphate, it inhibits RNA synthesis. How does cordycepin 5′-triphosphate block the synthesis of RNA?

37. *Silent RNA.* The code word GGG cannot be deciphered in the same way as can UUU, CCC, and AAA, because poly(G) does not act as a template. Poly(G) forms a triple-stranded helical structure. Why is it an ineffective template?

38. *Sometimes it is not so bad.* What is meant by the degeneracy of the genetic code?

39. *In fact, it can be good.* What is the biological benefit of a degenerate genetic code?

40. *To bring together as associates.* Match the components in the right-hand column with the appropriate process in the left-hand column.

(a) Replication ______
(b) Transcription ______
(c) Translation ______

1. RNA polymerase
2. DNA polymerase
3. Ribosome
4. dNTP
5. tRNA
6. NTP
7. mRNA
8. primer
9. rRNA
10. promoter

41. *A lively contest.* Match the components in the right-hand column with the appropriate process in the left-hand column.

(a) fMet ______
(b) Shine–Dalgarno ______
(c) intron ______
(d) exon ______
(e) pre-mRNA ______
(f) mRNA ______
(g) spliceosome ______

1. continuous message
2. removed during processing
3. the first of many amino acids
4. joins exons
5. joined to make the final message
6. locates the start site
7. discontinuous message

42. *Two from one.* Synthetic RNA molecules of defined sequence were instrumental in deciphering the genetic code. Their synthesis first required the synthesis of DNA molecules to serve as templates. Har Gobind Khorana synthesized, by organic-chemical methods, two complementary deoxyribonucleotides, each with nine residues: $d(TAC)_3$ and $d(GTA)_3$. Partly overlapping duplexes that formed on mixing these oligonucleotides then served as templates for the synthesis by DNA polymerase of long, repeating double-helical DNA strands. The next step was to obtain long polyribonucleotide chains with a sequence complementary to only one of the two DNA strands. How did Khorana obtain only poly(UAC)? Only poly(GUA)?

43. *Triple entendre.* The RNA transcript of a region of T4 phage DNA contains the sequence 5′-AAAUGAGGA-3′. This sequence encodes three different polypeptides. What are they?

44. *A new translation.* A transfer RNA with a UGU anticodon is enzymatically conjugated to ^{14}C-labeled cysteine. The cysteine unit is then chemically modified to alanine (with the use of Raney nickel, which removes the sulfur atom of

cysteine). The altered aminoacyl-tRNA is added to a protein-synthesizing system containing normal components except for this tRNA. The mRNA added to this mixture contains the following sequence:

5′–UUUUGCCAUGUUUGUGCU–3′

What is the sequence of the corresponding radiolabeled peptide?

45. *A tricky exchange*. Define exon shuffling and explain why its occurrence might be an evolutionary advantage.

46. *From one, many*. Explain why alternative splicing increases the coding capacity of the genome.

47. *The unity of life*. What is the significance of the fact that human mRNA can be accurately translated in *E. coli*?

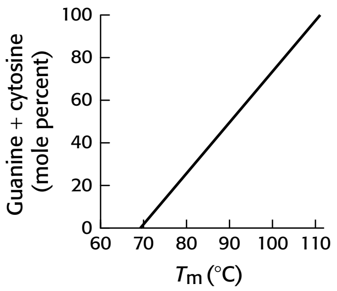
Chapter Integration Problems

48. *Back to the bench*. A protein chemist told a molecular geneticist that he had found a new mutant hemoglobin in which aspartate replaced lysine. The molecular geneticist expressed surprise and sent his friend scurrying back to the laboratory. (a) Why did the molecular geneticist doubt the reported amino acid substitution? (b) Which amino acid substitutions would have been more palatable to the molecular geneticist?

49. *Eons apart*. The amino acid sequences of a yeast protein and a human protein having the same function are found to be 60% identical. However, the corresponding DNA sequences are only 45% identical. Account for this differing degree of identity.

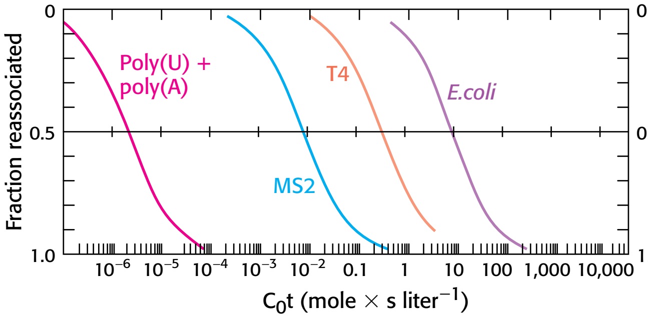
Data Interpretation Problems

50. *3 is greater than 2*. The adjoining illustration graphs the relation between the percentage of GC base pairs in DNA and the melting temperature. Suggest a plausible explanation for these results.



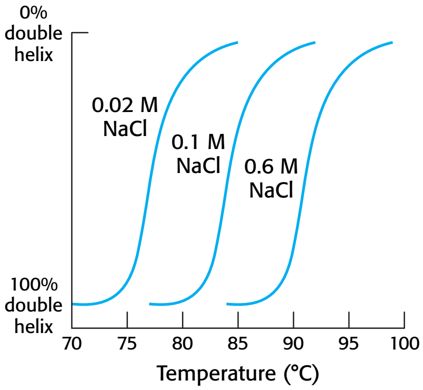
[Data from R. J. Britten and D. E. Kohne, *Science* 161:529–540, 1968.]

51. *Blast from the past*. The illustration below is a graph called a C_0t curve (pronounced "cot"). The *y*-axis shows the percentage of DNA that is double stranded. The *x*-axis is the product of the concentration of DNA and the time required for the double-stranded molecules to form. Explain why the mixture of poly(A) and poly(U) and the three DNAs shown vary in the C_0t value required to completely anneal. MS2 and T4 are bacterial viruses (bacteriophages) with genome sizes of 3569 and 168,903 bp, respectively. The *E. coli* genome is 4.6 × 10^6 bp.



[Data from J. Marmur and P. Doty, *J. Mol. Biol.* 5:120, 1962.]

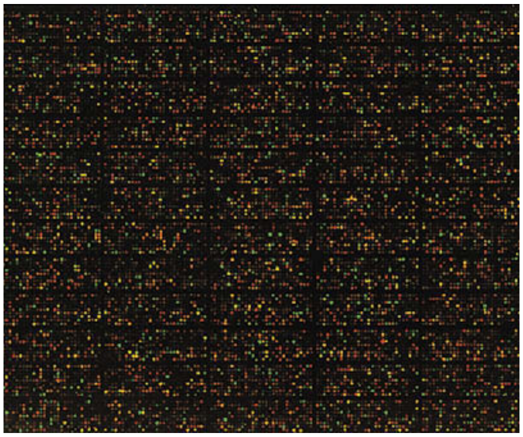
52. *Salt to taste*. The graph below shows the effect of salt concentration on melting temperature of bacterial DNA. How does salt concentration affect the melting temperature of DNA? Account for this effect.



CHAPTER

5

Exploring Genes and Genomes



Processes such as the development from a caterpillar into a butterfly entail dramatic changes in patterns of gene expression. The expression levels of thousands of genes can be monitored through the use of DNA arrays. At the right, a DNA microarray reveals the expression levels of more than 12,000 human genes; the brightness and color of each spot indicates a change in the expression level of the corresponding gene. [(Left) Cathy Keifer/istockphoto.com. (Right) Agilent Technologies.]

OUTLINE

5.1 The Exploration of Genes Relies on Key Tools

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

5.3 Complete Genomes Have Been Sequenced and Analyzed

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

Since its emergence in the 1970s, recombinant DNA technology has revolutionized biochemistry. The genetic endowment of organisms can now be precisely changed in designed ways. Recombinant DNA technology is the fruit of several decades of basic research on DNA, RNA, and viruses. It depends, first, on having enzymes that can cut, join, and replicate DNA and those that can reverse transcribe RNA. Restriction enzymes cut very long DNA molecules into specific fragments that can be manipulated; DNA ligases join the fragments together. Many kinds of restriction enzymes are available. By applying this assortment cleverly, researchers can treat DNA sequences as modules that can be moved at will from one DNA molecule to another. Thus, recombinant DNA technology is based on the use of enzymes that act on nucleic acids as substrates.

A second foundation is the base-pairing language that allows complementary sequences to recognize and bind to each other. Hybridization with complementary DNA (cDNA) or RNA probes is a sensitive means of detecting specific nucleotide sequences. In recombinant DNA technology, base-pairing is used to construct new combinations of DNA as well as to detect and amplify particular sequences.

Third, powerful methods have been developed for determining the sequence of nucleotides in DNA. These methods have been harnessed to sequence complete genomes: first, small genomes from viruses; then, larger genomes from bacteria; and, finally, eukaryotic genomes, including the 3-billion-base-pair human genome. Scientists are just beginning to exploit the enormous information content of these genome sequences.

Finally, recombinant DNA technology critically depends on our ability to deliver foreign DNA into host organisms. For example, DNA fragments can be inserted into plasmids, where they can be replicated within a short period of time in their bacterial hosts. In addition, viruses efficiently deliver their own DNA (or RNA) into hosts, subverting them either to replicate the viral genome and produce viral proteins or to incorporate viral DNA into the host genome.

These new methods have wide-ranging benefits across a broad spectrum of disciplines, including biotechnology, agriculture, and medicine. Among these benefits is the dramatic expansion of our understanding of human disease. Throughout this chapter, a specific disorder, amyotrophic lateral sclerosis (ALS), will be used to illustrate the effect that recombinant DNA technology has had on our knowledge of disease mechanisms. ALS was first described clinically in 1869 by the French neurologist Jean-Martin Charcot as a fatal neurodegenerative disease of progressive weakening and atrophy of voluntary muscles. ALS is commonly referred to as Lou Gehrig's Disease, for the baseball legend whose career and life were prematurely cut short as a result of this devastating disease. For many years, little progress had been made in the study of the mechanisms underlying ALS. As we shall see, significant advances have been made with the use of research tools facilitated by recombinant DNA technology.

5.1 The Exploration of Genes Relies on Key Tools

The rapid progress in biotechnology—indeed its very existence—is a result of a few key techniques.

1. *Restriction-enzyme analysis.* Restriction enzymes are precise molecular scalpels that allow an investigator to manipulate DNA segments.

2. *Blotting techniques.* Southern and northern blots are used to separate and identify DNA and RNA sequences, respectively. The western blot, which uses antibodies to characterize proteins, was described in Chapter 3.

3. *DNA sequencing.* The precise nucleotide sequence of a molecule of DNA can be determined. Sequencing has yielded a wealth of information concerning gene architecture, the control of gene expression, and protein structure.

4. *Solid-phase synthesis of nucleic acids.* Precise sequences of nucleic acids can be synthesized de novo and used to identify or amplify other nucleic acids.

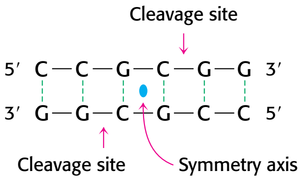
5. *The polymerase chain reaction* (PCR). The polymerase chain reaction leads to a billionfold amplification of a segment of DNA. One molecule of DNA can be amplified to quantities that permit characterization and manipulation. This powerful technique can be used to detect pathogens and genetic diseases, determine the source of a hair left at the scene of a crime, and resurrect genes from the fossils of extinct organisms.

A final set of techniques relies on the computer, without which, it would be impossible to catalog, access, and characterize the abundant information generated by the methods outlined above. Such uses of the computer will be presented in Chapter 6.

Restriction enzymes split DNA into specific fragments

Restriction enzymes, also called *restriction endonucleases*, recognize specific base sequences in double-helical DNA and cleave both strands of that duplex at specific places. To biochemists, these exquisitely precise scalpels are marvelous gifts of nature. They are indispensable for analyzing chromosome structure, sequencing very long DNA molecules, isolating genes, and creating new DNA molecules that can be cloned.

Restriction enzymes are found in a wide variety of prokaryotes. Their biological role is to cleave foreign DNA molecules, providing the host organism with a primitive immune system. Many restriction enzymes recognize specific sequences of four to eight base pairs and hydrolyze a phosphodiester bond in each strand in this region. A striking characteristic of these cleavage sites is that they almost always possess *twofold rotational symmetry.* In other words, the recognized sequence is *palindromic,* or an inverted repeat, and the cleavage sites are symmetrically positioned. For example, the sequence recognized by a restriction enzyme from *Streptomyces achromogenes* is



Palindrome
A word, sentence, or verse that reads the same from right to left as it does from left to right.

Radar
Senile felines
Do geese see God?
Roma tibi subito motibus ibit amor

Derived from the Greek *palindromos*, "running back again."

In each strand, the enzyme cleaves the C–G phosphodiester bond on the 3′ side of the symmetry axis. As we shall see in Chapter 9, this symmetry corresponds to that of the structures of the restriction enzymes themselves.

Several hundred restriction enzymes have been purified and characterized. Their names consist of a three-letter abbreviation for the host organism (e.g., Eco for *Escherichia coli,* Hin for *Haemophilus influenzae,* Hae for *Haemophilus aegyptius*) followed by a strain designation (if needed) and a roman numeral (to distinguish multiple enzymes from the same strain). The specificities of several of these enzymes are shown in Figure 5.1.

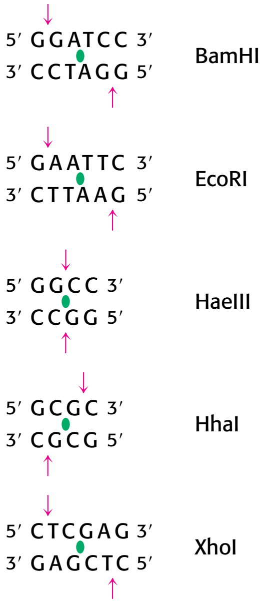


FIGURE 5.1 Specificities of some restriction endonucleases. The sequences that are recognized by these enzymes contain a twofold axis of symmetry. The two strands in these regions are related by a 180-degree rotation about the axis marked by the green symbol. The cleavage sites are denoted by red arrows. The abbreviated name of each restriction enzyme is given at the right of the sequence that it recognizes. Note that the cuts may be staggered or even.

Restriction enzymes are used to cleave DNA molecules into specific fragments that are more readily analyzed and manipulated than the entire parent molecule. For example, the 5.1-kb circular duplex DNA of the tumor-producing SV40 virus is cleaved at one site by EcoRI, at four sites by HpaI, and at 11 sites by HindIII. A piece of DNA, called a restriction fragment, produced by the action of one restriction enzyme can be specifically cleaved into smaller fragments by another restriction enzyme. The pattern of such fragments can serve as a *fingerprint* of a DNA molecule, as will be considered shortly. Indeed, complex chromosomes containing hundreds of millions of base pairs can be mapped by using a series of restriction enzymes.

Restriction fragments can be separated by gel electrophoresis and visualized

In Chapter 3, we considered the use of gel electrophoresis to separate protein molecules (Section 3.1). Because the phosphodiester backbone of DNA is highly negatively charged, this technique is also suitable for the separation of nucleic acid fragments. Among the many applications of DNA electrophoresis are the detection of mutations that affect restriction fragment size (such as insertions and deletions) and the isolation, purification, and quantitation of a specific DNA fragment.

For most gels, the shorter the DNA fragment, the farther the migration. Polyacrylamide gels are used to separate, by size, fragments containing as many as 1000 base pairs, whereas more-porous agarose gels are used to resolve mixtures of larger fragments (as large as 20 kb). An important feature of these gels is their high resolving power. In certain kinds of gels, fragments differing in length by just one nucleotide out of several hundred can be distinguished. Bands or spots of radioactive DNA in gels can be visualized by autoradiography. Alternatively, a gel can be stained with a dye such as ethidium bromide, which fluoresces an intense orange under irradiation with ultraviolet light when bound to a double-helical DNA molecule (Figure 5.2). A band containing only 10 ng of DNA can be readily seen.

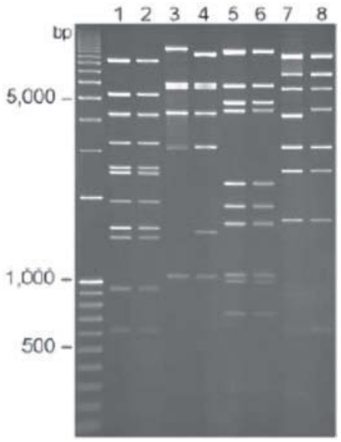


FIGURE 5.2 Gel-electrophoresis pattern of a restriction digest. This gel shows the fragments produced by cleaving DNA from two viral strains (odd- vs. even-numbered lanes) with each of four restriction enzymes. These fragments were made fluorescent by staining the gel with ethidium bromide. [Data from Carr et al., Emerging Infectious Diseases, www.cdc.gov/eid, 17(8), August 2011.]

It is often necessary to determine if a particular base sequence is represented in a given DNA sample. For example, one may wish to confirm the presence of a specific mutation in genomic DNA isolated from patients known to be at risk for a particular disease. This specific sequence can be identified by hybridizing it with a labeled complementary DNA strand (Figure 5.3). A mixture of restriction fragments is separated by electrophoresis through an agarose gel, denatured to form single-stranded DNA, and transferred to a nitrocellulose sheet. The positions of the DNA fragments in the gel are preserved during the transfer. The nitrocellulose is then exposed to a ^{32}P-labeled or fluorescently tagged *DNA probe*, a short stretch of single-stranded DNA which contains a known base sequence. The probe hybridizes with a restriction fragment having a complementary sequence, and autoradiography or fluorescence imaging then reveals the position of the restriction-fragment–probe duplex. A particular fragment amid a million others can be readily identified in this way. This powerful technique is named *Southern blotting,* for its inventor Edwin Southern.

In a similar manner, RNA molecules of a specific sequence can also be readily identified. After separation by gel electrophoresis and transfer to nitrocellulose, specific sequences can be detected by DNA probes. This analogous technique for the analysis of RNA has been whimsically termed *northern blotting*. A further play on words accounts for the term *western blotting,* which refers to a technique for detecting a particular protein by staining with specific antibody (Section 3.3).

DNA can be sequenced by controlled termination of replication

The analysis of DNA structure and its role in gene expression have been markedly facilitated by the development of powerful techniques for the *sequencing* of DNA molecules. One of the first and most widely-used techniques for DNA sequencing is *controlled termination of replication,* also referred to as the Sanger dideoxy method for its pioneer, Frederick Sanger. The key to this approach is

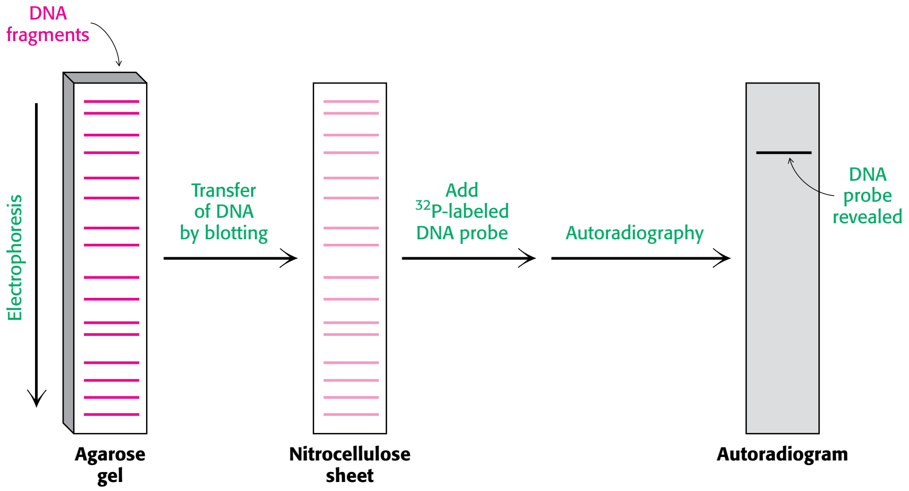
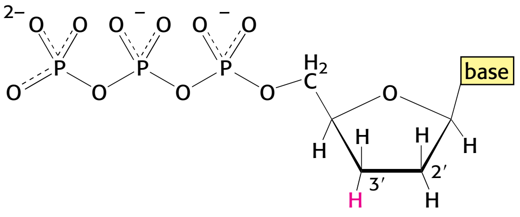


FIGURE 5.3 Southern blotting. A DNA fragment containing a specific sequence can be identified by separating a mixture of fragments by electrophoresis, transferring them to nitrocellulose, and hybridizing with a ^{32}P-labeled probe complementary to the sequence. The fragment containing the sequence is then visualized by autoradiography.

the generation of DNA fragments whose length is determined by the last base in the sequence (Figure 5.4). In the current application of this method, a DNA polymerase is used to make the complement of a particular sequence within a single-stranded DNA molecule. The synthesis is primed by a chemically synthesized fragment that is complementary to a part of the sequence known from other studies. In addition to the four deoxyribonucleoside triphosphates, the reaction mixture contains a small amount of the 2′,3′-*dideoxy analog* of each nucleotide, each carrying a different fluorescent label attached to the base (e.g., a green emitter for termination at A and a red one for termination at T).


2−
O
O
O
−
O
O
−
O
P
P
P
O
O
O
O
H2
C
O
base
H
H
H
H
3′
2′
H
H

2′, 3′-Dideoxy analog

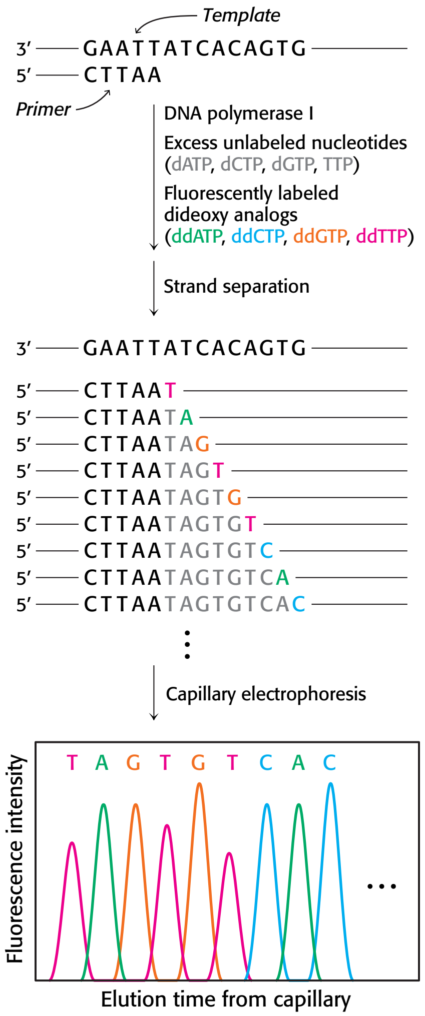

Template
3′ GAATTATCACAGTG
5′ CTTAA
Primer
DNA polymerase I
Excess unlabeled nucleotides (dATP, dCTP, dGTP, TTP)
Fluorescently labeled dideoxy analogs (ddATP, ddCTP, ddGTP, ddTTP)
Strand separation
3′ GAATTATCACAGTG
5′ CTTAAT
5′ CTTAATA
5′ CTTAATAG
5′ CTTAATAGT
5′ CTTAATAGTG
5′ CTTAATAGTGT
5′ CTTAATAGTGTC
5′ CTTAATAGTGTCA
5′ CTTAATAGTGTCAC
Capillary electrophoresis
T A G T G T C A C
Fluorescence intensity
Elution time from capillary

FIGURE 5.4 Fluorescence detection of oligonucleotide fragments produced by the dideoxy method. A sequencing reaction is performed with four chain-terminating dideoxy nucleotides, each labeled with a tag that fluoresces at a different wavelength. The color of each fragment indicates the identity of the last base in the chain. The fragments are separated by size using capillary electrophoresis and the fluorescence at each of the four wavelengths indicates the sequence of the complement of the original DNA template.

The incorporation of this analog blocks further growth of the new chain because it lacks the 3′-hydroxyl terminus needed to form the next phosphodiester bond. The concentration of the dideoxy analog is low enough that chain termination will take place only occasionally. The polymerase will insert the correct nucleotide sometimes and the dideoxy analog other times, stopping the reaction. For instance, if the dideoxy analog of dATP is present, fragments of various lengths are produced, but all will be terminated by the dideoxy analog. Importantly, this dideoxy analog of dATP will be inserted only where a T was located in the DNA being sequenced. Thus, the fragments of different length will correspond to the positions of T.

The resulting fragments are separated by a technique known as *capillary electrophoresis,* in which the mixture is passed through a very narrow tube containing a gel matrix at high voltage to achieve efficient separation within a short time. As the DNA fragments emerge from the capillary, they are detected by their fluorescence; the sequence of their colors directly gives the base sequence. Sequences of as many as 1000 bases can be determined in this way. Indeed, automated Sanger sequencing machines can read more than 1 million bases per day.

DNA probes and genes can be synthesized by automated solid-phase methods

DNA strands, like polypeptides (Section 3.4), can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to a solid support. The activated monomers are protected *deoxyribonucleoside 3′-phosphoramidites.* In step 1, the 3′-phosphorus atom of this incoming unit becomes joined to the 5′-oxygen atom of the growing chain to form a *phosphite triester* (Figure 5.5). The 5′-OH group of the activated monomer is unreactive because it is blocked by a dimethoxytrityl (DMT) protecting group, and the 3′-phosphoryl oxygen atom is rendered unreactive by attachment of the β-cyanoethyl (βCE) group. Likewise, amino groups on the purine and pyrimidine bases are blocked.

Coupling is carried out under anhydrous conditions because water reacts with phosphoramidites. In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent). In step 3, the DMT protecting group on the 5′-OH group of the growing chain is removed by the addition of dichloroacetic acid, which

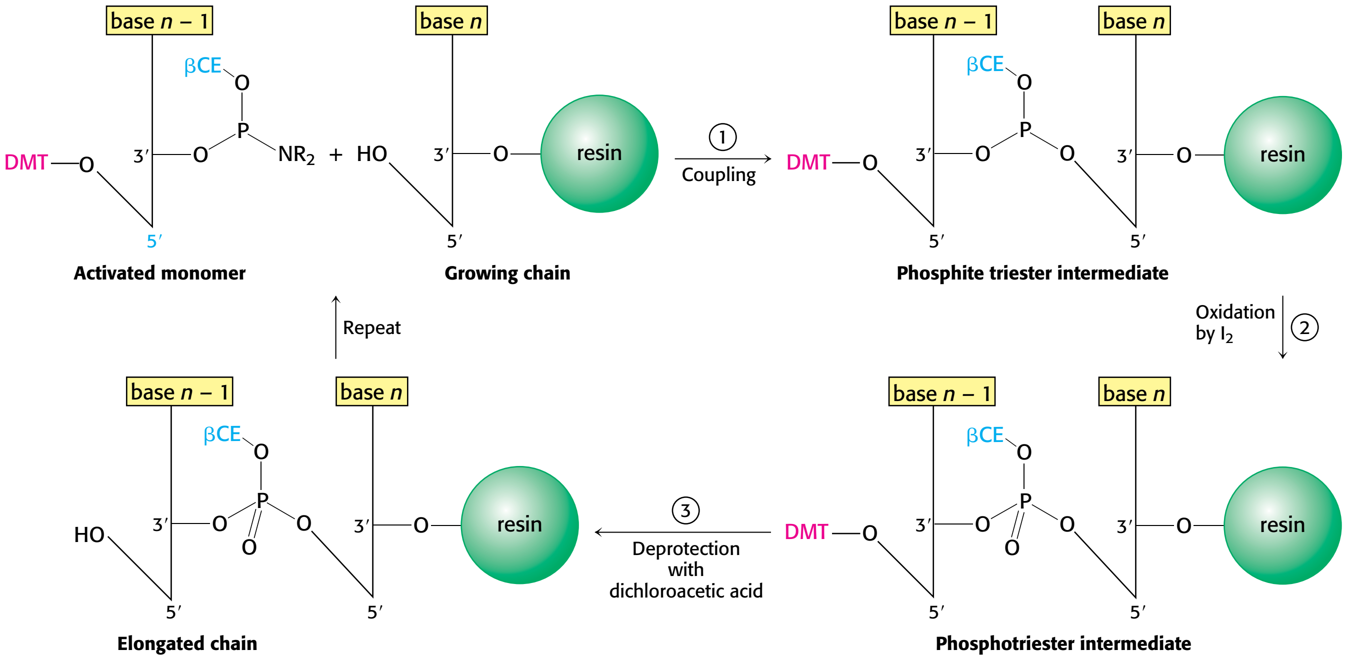
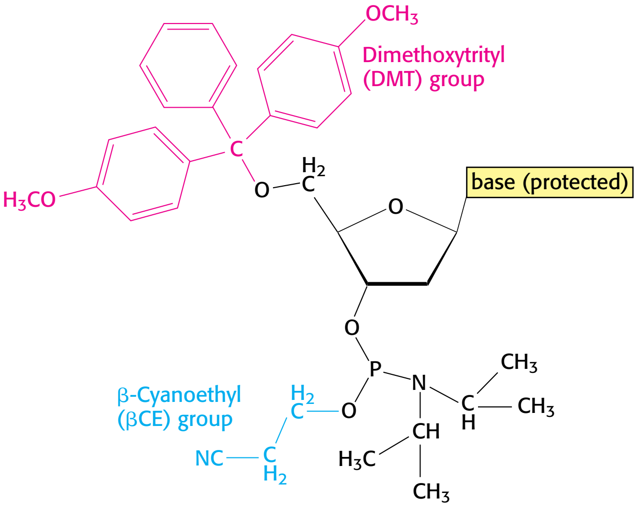

base n – 1
base n
βCE
O
P
DMT
O
3′
NR₂
+ HO
resin
5′
Activated monomer
Growing chain
① Coupling
Phosphite triester intermediate
Oxidation by I₂ ②
Repeat
Elongated chain
③ Deprotection with dichloroacetic acid
Phosphotriester intermediate

FIGURE 5.5 Solid-phase synthesis of a DNA chain by the phosphite triester method. The activated monomer added to the growing chain is a deoxyribonucleoside 3′-phosphoramidite containing a dimethoxytrityl (DMT) protecting group on its 5′-oxygen atom, a β-cyanoethyl (βCE) protecting group on its 3′-phosphoryl oxygen atom, and a protecting group on the base.


OCH_3
Dimethoxytrityl (DMT) group
H_3CO
C
O
H_2
base (protected)
β-Cyanoethyl (βCE) group
NC
P
N
CH_3
CH
H_3C

A deoxyribonucleoside 3′-phosphoramidite with DMT and βCE attached

leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition. Each cycle takes only about 10 minutes and usually elongates more than 99% of the chains.

This solid-phase approach is ideal for the synthesis of DNA, as it is for polypeptides, because the desired product stays on the insoluble support until the final release step. All the reactions take place in a single vessel, and excess soluble reagents can be added to drive reactions to completion. At the end of each step, soluble reagents and by-products are washed away from the resin that bears the growing chains. At the end of the synthesis, NH_3 is added to remove all protecting groups and release the oligonucleotide from the solid support. Because elongation is never 100% complete, the new DNA chains are of diverse lengths—the desired chain is the longest one. The sample can be purified by high-performance liquid chromatography or by electrophoresis on polyacrylamide gels. DNA chains of as many as 100 nucleotides can be readily synthesized by this automated method.

The ability to rapidly synthesize DNA chains of any selected sequence opens many experimental avenues. For example, a synthesized oligonucleotide labeled at one end with ^{32}P or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes. The use of labeled oligonucleotides as DNA probes is powerful and general. For example, a DNA probe that can base-pair to a known complementary sequence in a chromosome can serve as the starting point of an exploration of adjacent uncharted DNA. Such a probe can be used as a *primer* to initiate the replication of neighboring DNA by DNA polymerase. An exciting application of the solid-phase approach is the synthesis of new tailor-made genes. New proteins with novel properties can now be produced in abundance by the expression of synthetic genes. Finally, the synthetic scheme heretofore described can be slightly modified for the solid-phase synthesis of RNA oligonucleotides, which can be very powerful reagents for the degradation of specific mRNA molecules in living cells by a technique known as RNA interference (Section 5.4).

Selected DNA sequences can be greatly amplified by the polymerase chain reaction

In 1984, Kary Mullis devised an ingenious method called the *polymerase chain reaction* (PCR) for amplifying specific DNA sequences. Consider a DNA duplex consisting of a target sequence surrounded by nontarget DNA. Millions of copies of the target sequences can be readily obtained by PCR if the sequences flanking the target are known. PCR is carried out by adding the following components to a solution containing the target sequence: (1) a pair of primers that hybridize with the flanking sequences of the target, (2) all four deoxyribonucleoside triphosphates (dNTPs), and (3) a heat-stable DNA polymerase. A PCR cycle consists of three steps (Figure 5.6).

1. *Strand separation.* The two strands of the parent DNA molecule are separated by heating the solution to 95°C for 15 s.

2. *Hybridization of primers.* The solution is then abruptly cooled to 54°C to allow each primer to hybridize to a DNA strand. One primer hybridizes to the 3′ end of the target on one strand, and the other primer hybridizes to the 3′ end on the complementary target strand. Parent DNA duplexes do not form, because the primers are present in large excess. Primers are typically from 20 to 30 nucleotides long.

3. *DNA synthesis.* The solution is then heated to 72°C, the optimal temperature for heat-stable polymerases. One such enzyme is *Taq* DNA polymerase, which is derived from *Thermus aquaticus,* a thermophilic bacterium that lives in hot springs. The polymerase elongates both primers in the direction of the target sequence because DNA synthesis is in the 5′-to-3′ direction. DNA synthesis takes place on both strands but extends beyond the target sequence.

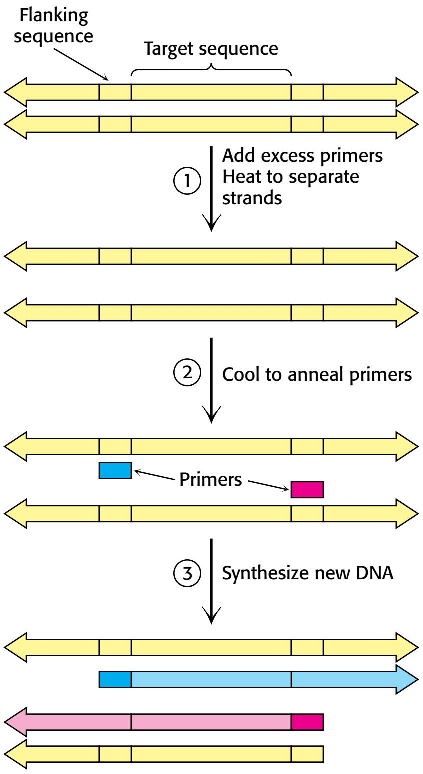


FIGURE 5.6 The first cycle in the polymerase chain reaction (PCR). A cycle consists of three steps: DNA double strand separation, the hybridization of primers, and the extension of primers by DNA synthesis.

These three steps—strand separation, hybridization of primers, and DNA synthesis—constitute one cycle of the PCR amplification and can be carried out repetitively just by changing the temperature of the reaction mixture. The thermostability of the polymerase makes it feasible to carry out PCR in a closed container; no reagents are added after the first cycle. At the completion of the second cycle, four duplexes containing the target sequence have been generated (Figure 5.7). Of the eight DNA strands comprising these duplexes, two short strands constitute only the target sequence—the sequence including and bounded by the primers. Subsequent cycles will amplify the target sequence exponentially. Ideally, after n cycles, the desired sequence is amplified 2^n-fold. The amplification is a millionfold after 20 cycles and a billionfold after 30 cycles, which can be carried out in less than an hour.

Several features of this remarkable method for amplifying DNA are noteworthy. First, the sequence of the target need not be known. All that is required is knowledge of the flanking sequences so that complementary primers can be synthesized. Second, the target can be much larger than the primers. Targets larger than 10 kb have been amplified by PCR. Third, primers do not have to be perfectly matched to flanking sequences to amplify targets. With the use of primers derived from a gene of known sequence, it is possible to search for variations on the theme. In this way, families of genes are being discovered by PCR. Fourth, PCR is highly specific because of the

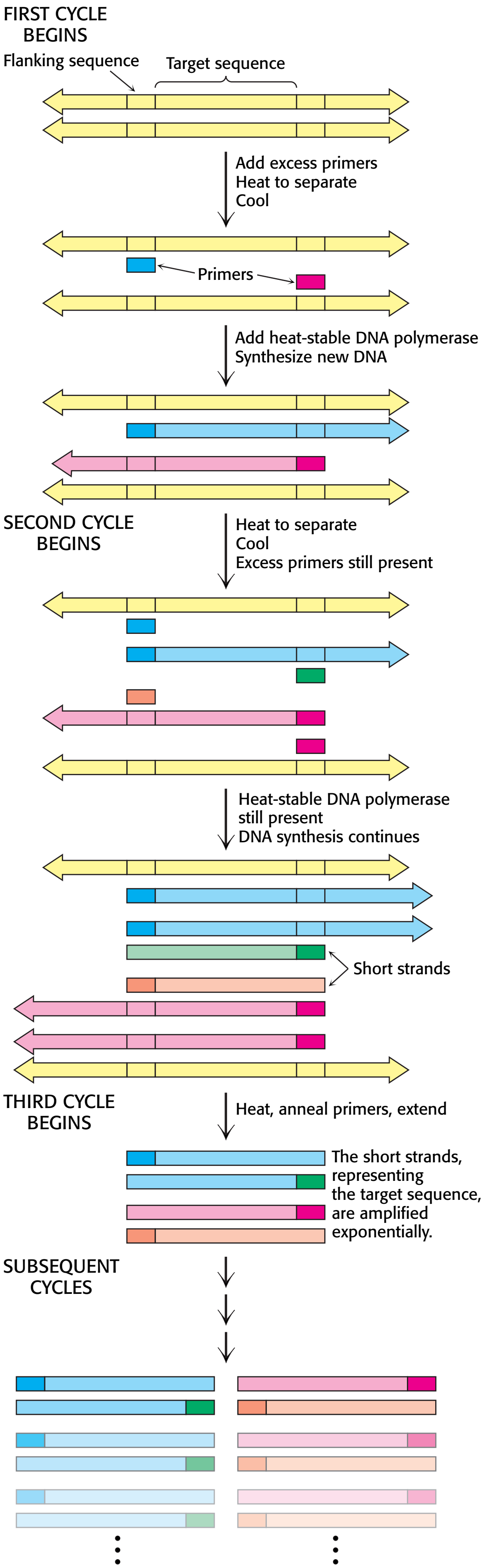

FIRST CYCLE BEGINS
Flanking sequence
Target sequence
Add excess primers
Heat to separate
Cool
Primers
Add heat-stable DNA polymerase
Synthesize new DNA
SECOND CYCLE BEGINS
Heat to separate
Cool
Excess primers still present
Heat-stable DNA polymerase still present
DNA synthesis continues
Short strands
THIRD CYCLE BEGINS
Heat, anneal primers, extend
The short strands, representing the target sequence, are amplified exponentially.
SUBSEQUENT CYCLES

FIGURE 5.7 Multiple cycles of the polymerase chain reaction. The two short strands produced at the end of the third cycle (along with longer stands not shown) represent the target sequence. Subsequent cycles will amplify the target sequence exponentially and the parent sequence arithmetically.

stringency of hybridization at relatively high temperature. *Stringency* is the required closeness of the match between primer and target, which can be controlled by temperature and salt. At high temperatures, only the DNA between hybridized primers is amplified. A gene constituting less than a millionth of the total DNA of a higher organism is accessible by PCR. Fifth, PCR is exquisitely sensitive. A single DNA molecule can be amplified and detected.

PCR is a powerful technique in medical diagnostics, forensics, and studies of molecular evolution

PCR can provide valuable diagnostic information in medicine. Bacteria and viruses can be readily detected with the use of specific primers. For example, PCR can reveal the presence of small amounts of DNA from the human immunodeficiency virus (HIV) in persons who have not yet mounted an immune response to this pathogen. In these patients, assays designed to detect antibodies against the virus would yield a false negative test result. Finding *Mycobacterium tuberculosis* bacilli in tissue specimens is slow and laborious. With PCR, as few as 10 tubercle bacilli per million human cells can be readily detected. PCR is a promising method for the early detection of certain cancers. This technique can identify mutations of certain growth-control genes, such as the *ras* genes (Chapter 14). The capacity to greatly amplify selected regions of DNA can also be highly informative in monitoring cancer chemotherapy. Tests using PCR can detect when cancerous cells have been eliminated and treatment can be stopped; they can also detect a relapse and the need to immediately resume treatment. PCR is ideal for detecting leukemias caused by chromosomal rearrangements.

In addition, PCR has made an impact on forensics and legal medicine. An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population. For example, variations at one specific location determine a person's HLA type (human leukocyte antigen type; Section 34.5); organ transplants are rejected when the HLA types of the donor and recipient are not sufficiently matched. PCR amplification of multiple genes is being used to establish biological parentage in disputed paternity and immigration cases. Analyses of blood stains and semen samples by PCR have implicated guilt or innocence in numerous assault and rape cases (Figure 5.8). The root of a single shed hair found at a crime scene contains enough DNA for typing by PCR.

DNA is a remarkably stable molecule, particularly when shielded from air, light, and water. Under such circumstances, large fragments of DNA can remain intact for thousands of years or longer. PCR provides an ideal method for amplifying such ancient DNA molecules so that they can be detected and characterized (Section 6.5). PCR can also be used to amplify DNA from microorganisms that have not yet been isolated and cultured. As will be discussed in Chapter 6, sequences from these PCR products can be sources of considerable insight into evolutionary relationships between organisms.

The tools for recombinant DNA technology have been used to identify disease-causing mutations

Let us consider how the techniques just described have been utilized in concert to study ALS, introduced at the beginning of this chapter. Five percent of all patients suffering from ALS have family members who also have been diagnosed with the disease. A heritable disease pattern is indicative of a strong genetic component of disease causation. To identify these disease-causing genetic alterations, researchers identify *polymorphisms* (instances of genetic variation) within an affected family that correlate with the emergence of disease. These polymorphisms may themselves cause disease or be closely linked to the responsible genetic alteration. *Restriction-fragment-length polymorphisms* (RFLPs) are polymorphisms within restriction sites that change the sizes of DNA fragments produced by the appropriate restriction enzyme. Using restriction digests and Southern blots of the DNA from members of ALS-affected families, researchers identified RFLPs that were found preferentially in those family members diagnosed with the disease. For some of these families, strong evidence was obtained for the disease-causing mutation within a specific region of chromosome 21.

After the probable location of one disease-causing gene had been identified, this same research group compared the locations of the ALS-associated RFLPs with the known sequence of chromosome 21. They noted that this chromosomal locus contains the *SOD1* gene, which encodes the Cu/Zn superoxide dismutase protein SOD1, an enzyme important for the protection of cells against oxidative damage (Section 18.3). PCR amplification of regions of the *SOD1* gene from the DNA of affected family members, followed by Sanger dideoxy sequencing of the targeted fragment, enabled the identification of 11 disease-causing mutations from 13 different families. This work was pivotal for focusing further inquiry into the roles that superoxide dismutase and its corresponding mutant forms play in the pathology of some forms of ALS.

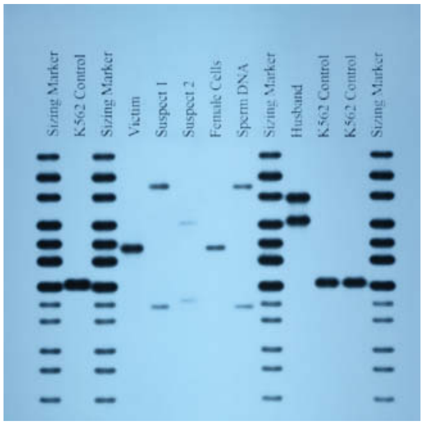


FIGURE 5.8 DNA and forensics. DNA isolated from sperm obtained during the examination of a rape victim was amplified by PCR, then compared with DNA from the victim and three potential suspects—the victim's husband and two additional individuals—using gel electrophoresis and autoradiography. Sperm DNA matched the pattern of Suspect 1, but not that of Suspect 2 or the victim's husband. Sizing marker and K562 lanes refer to control DNA samples. [Martin Shields/Science Source.]

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

The development of recombinant DNA technology has taken biology from an exclusively analytical science to a synthetic one. New combinations of unrelated genes can be constructed in the laboratory by applying recombinant DNA techniques. These novel combinations can be cloned—amplified many-fold—by introducing them into suitable cells, where they are replicated by the DNA-synthesizing machinery of the host. The inserted genes are often transcribed and translated in their new setting. What is most striking is that the genetic endowment of the host can be permanently altered in a designed way.

Restriction enzymes and DNA ligase are key tools in forming recombinant DNA molecules

Let us begin by seeing how novel DNA molecules can be constructed in the laboratory. An essential tool for the manipulation of recombinant DNA is a *vector,* a DNA molecule that can replicate autonomously in an appropriate host organism. Vectors are designed to enable the rapid, covalent insertion of DNA fragments of interest. *Plasmids* (naturally occurring circles of DNA that act as accessory chromosomes in bacteria) and bacteriophage lambda (λ phage), a virus, are choice vectors for cloning in *E. coli*. The vector can be prepared for accepting a new DNA fragment by cleaving it at a single specific site with a restriction enzyme. For example, the plasmid pSC101, a 9.9-kb double-helical circular DNA molecule, is split at a unique site by the EcoRI restriction enzyme. The staggered cuts made by this enzyme produce *complementary single-stranded ends*, which have specific affinity for each other and hence are known as *cohesive* or *sticky ends*. Any DNA fragment can be inserted into this plasmid if it has the same cohesive ends. Such a fragment can be extracted from a larger piece of DNA by using the same restriction enzyme as was used to open the plasmid DNA (Figure 5.9).

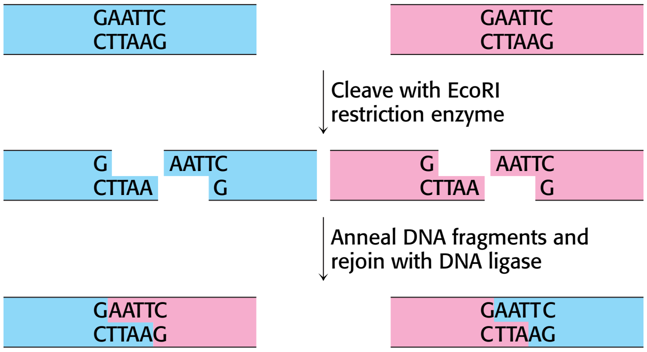

GAATTC
CTTAAG
GAATTC
CTTAAG
Cleave with EcoRI restriction enzyme
G
CTTAA
AATTC
G
G
CTTAA
AATTC
G
Anneal DNA fragments and rejoin with DNA ligase
GAATTC
CTTAAG
GAATTC
CTTAAG

FIGURE 5.9 Joining of DNA molecules by the cohesive-end method. Two DNA molecules, cleaved with a common restriction enzyme such as EcoRI, can be ligated to form recombinant molecules.

The single-stranded ends of the fragment are then complementary to those of the cut plasmid. The DNA fragment and the cut plasmid can be annealed and then joined by *DNA ligase,* which catalyzes the formation of a phosphodiester bond at a break in a DNA chain. DNA ligase requires a free 3′-hydroxyl group and a 5′-phosphoryl group. Furthermore, the chains joined by ligase must be in a double helix. An energy source such as ATP or NAD$^+$ is required for the joining reaction, as will be discussed in Chapter 28.

What if the target DNA is not naturally flanked by the appropriate restriction sites? How is the fragment cut and annealed to the vector? The cohesive-end method for joining DNA molecules can still be used in these cases by adding a *short, chemically synthesized DNA linker* that can be cleaved by restriction enzymes. First, the linker is covalently joined to the ends of a DNA fragment. For example, the 5′ ends of a decameric linker and a DNA molecule are phosphorylated by polynucleotide kinase and then joined by the ligase from T4 phage (Figure 5.10). This ligase can form a covalent bond between blunt-ended double-helical DNA molecules. Cohesive ends are produced when these terminal extensions are cut by an appropriate restriction enzyme. Thus, *cohesive ends corresponding to a particular restriction enzyme can be added to virtually any DNA molecule*. We see here the fruits of combining enzymatic and synthetic chemical approaches in crafting new DNA molecules.

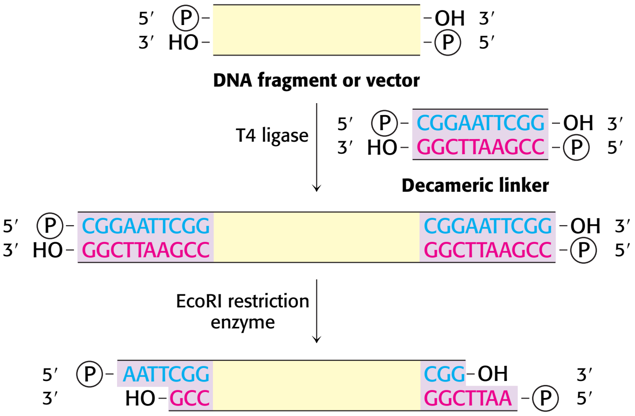


FIGURE 5.10 Formation of cohesive ends. Cohesive ends can be formed by the addition and cleavage of a chemically synthesized linker.

Plasmids and λ phage are choice vectors for DNA cloning in bacteria

Many plasmids and bacteriophages have been ingeniously modified by researchers to enhance the delivery of recombinant DNA molecules into bacteria and to facilitate the genetic selection of bacteria harboring these vectors. As already mentioned, plasmids are circular double-stranded DNA molecules that occur naturally in some bacteria. They range in size from two to several hundred kilobases. Plasmids carry genes for the inactivation of antibiotics, the production of toxins, and the breakdown of natural products. These accessory chromosomes can replicate independently of the host chromosome. In contrast with the host genome, they are dispensable under certain conditions. A bacterial cell may have no plasmids at all or it may house as many as 20 copies of a naturally-occurring plasmid.

Many plasmids have been optimized for a particular experimental task. Some engineered plasmids, for example, can achieve nearly a thousand copies per bacterial cell. One class of plasmids, known as *cloning vectors,* is particularly suitable for the facile insertion and replication of a collection of DNA fragments. These vectors often feature a *polylinker* region that includes many unique restriction sites within its sequence. This polylinker can be cleaved with many different restriction enzymes or combinations of enzymes, providing great versatility in the DNA fragments that can be inserted. In addition, these plasmids contain *reporter genes*, which encode rapidly-detectable markers such as antibiotic-resistance enzymes or fluorescent proteins. Creative placement of these reporter genes within these plasmids enables the rapid identification of those vectors that harbor the desired DNA insert. For example, in the cloning vector pUC18, insertion of DNA in the polylinker region (Figure 5.11) disrupts the *lacZ*α gene, an effect called *insertional inactivation*. This gene encodes an essential fragment of the protein *β-galactosidase*, an enzyme which naturally cleaves the milk sugar galactose (Section 11.2). β-Galactosidase also cleaves the synthetic substrate X-gal, releasing a blue dye. Bacterial cells containing a DNA insert at the polylinker will no longer produce the dye in the presence of X-gal, and are readily identified by their white color (Figure 5.12).

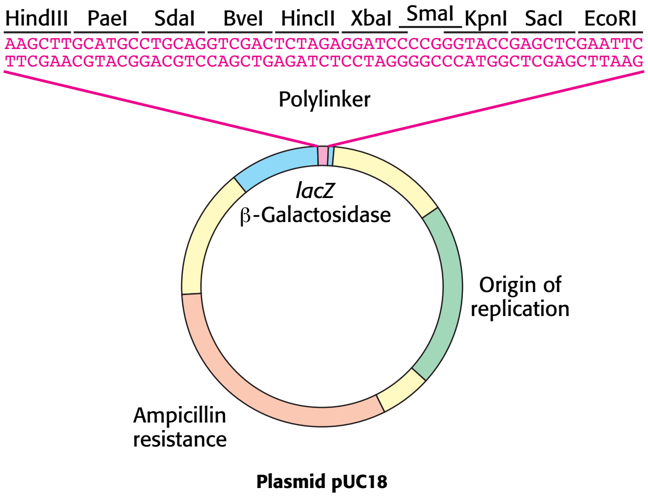

HindIII
PaeI
SdaI
BveI
HincII
XbaI
SmaI
KpnI
SacI
EcoRI
AAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCGAATTC
TTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGGGCCCATGGCTCGAGCTTAAG
Polylinker
lacZ
β-Galactosidase
Origin of replication
Ampicillin resistance
Plasmid pUC18

FIGURE 5.11 A polylinker in the plasmid pUC18. The plasmid pUC18 includes a polylinker within an essential fragment of the β-galactosidase gene (often called the *lacZ* gene). Insertion of a DNA fragment into one of the many restriction sites within this polylinker can be detected by the absence of β-galactosidase activity.

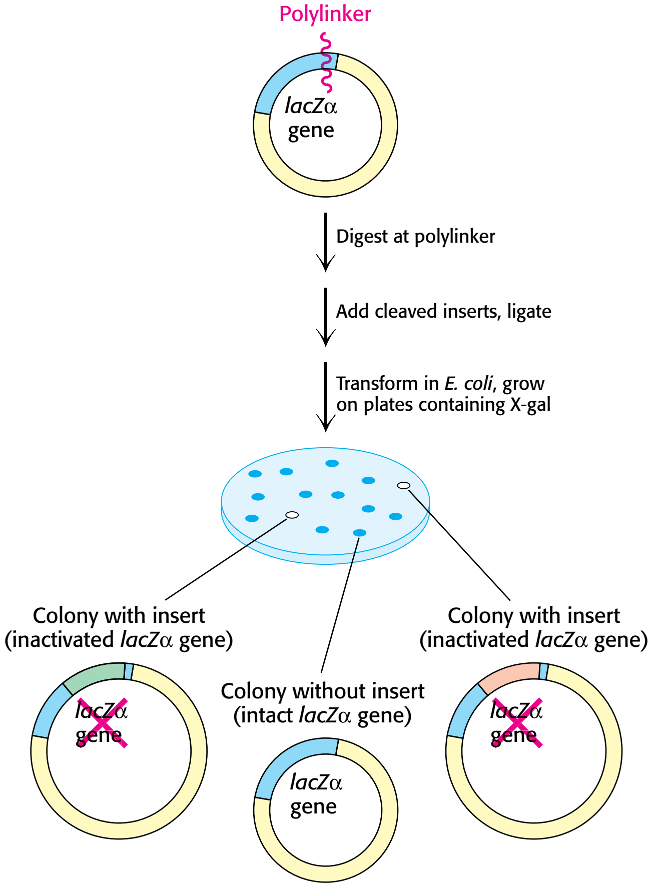

Polylinker
lacZα gene
Digest at polylinker
Add cleaved inserts, ligate
Transform in E. coli, grow on plates containing X-gal
Colony with insert (inactivated lacZα gene)
Colony without insert (intact lacZα gene)
Colony with insert (inactivated lacZα gene)
lacZα gene
lacZα gene
lacZα gene

FIGURE 5.12 Insertional inactivation. Successful insertion of DNA fragments into the polylinker region of pUC18 will result in the disruption of the β-galactosidase gene. Bacterial colonies that harbor such plasmids will no longer convert X-gal into a colored product, and will appear white on the plate.

Another class of plasmids has been optimized for use as *expression vectors* for the production of large amounts of protein. In addition to antibiotic-resistance genes, they contain promoter sequences designed to drive the transcription of large amounts of a protein-coding DNA sequence. In addition to polylinkers, these vectors often contain sequences flanking the cloning site that simplify the addition of fusion tags to the protein of interest (Section 3.1), greatly facilitating the purification of the overexpressed protein.

Another widely used vector, λ *phage,* enjoys a choice of life styles: this bacteriophage can destroy its host or it can become part of its host (Figure 5.13). In the *lytic pathway,* viral functions are fully expressed: viral DNA and proteins are quickly produced and packaged into virus particles, leading to the lysis (destruction) of the host cell and the sudden appearance of about 100 progeny virus particles, or *virions*. In the *lysogenic pathway,* the phage DNA becomes inserted into the host-cell genome and can be replicated together with host-cell DNA for many generations, remaining inactive. Certain environmental changes can trigger the expression of this dormant viral DNA, which leads to the formation of progeny viruses and lysis of the host. Large segments of the 48-kb DNA of λ phage are not essential for productive infection and can be replaced by foreign DNA, thus making λ phage an ideal vector.

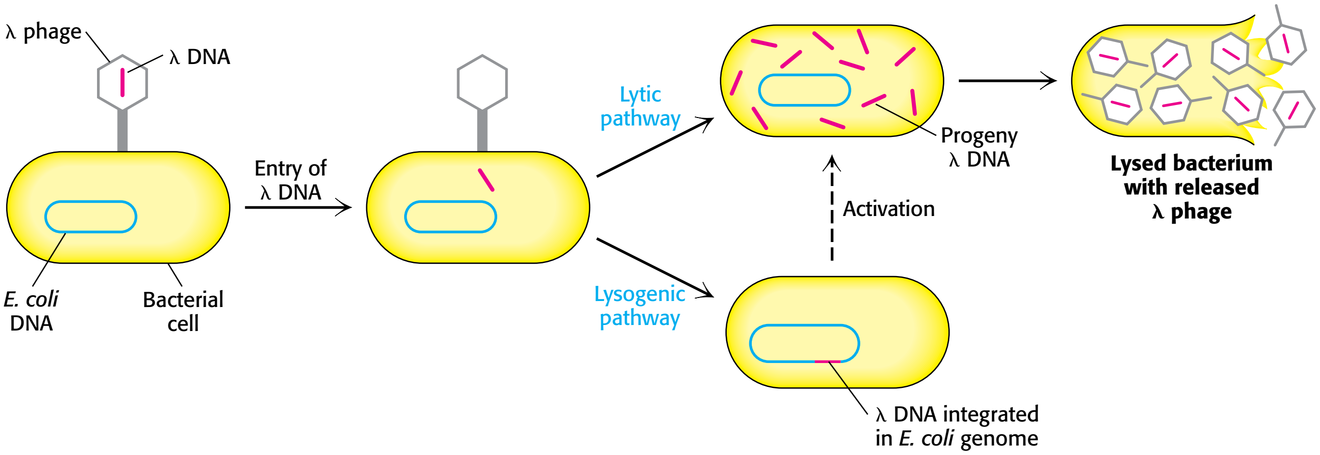


FIGURE 5.13 Alternative infection modes for λ phage. Lambda phage can multiply within a host and lyse it (lytic pathway) or its DNA can become integrated into the host genome (lysogenic pathway), where it is dormant until activated.

Mutant λ phages designed for cloning have been constructed. An especially useful one called λgt-λβ contains only two EcoRI cleavage sites instead of the five normally present (Figure 5.14). After cleavage, the middle segment of this λ DNA molecule can be removed. The two remaining pieces of DNA (called arms) have a combined length equal to 72% of a normal genome length. This amount of DNA is too little to be packaged into a λ particle, which can take up only DNA measuring from 78% to 105% of a normal genome. However, a suitably long DNA insert (such as 10 kb) between the two ends of λ DNA enables such a recombinant DNA molecule (93% of normal length) to be packaged. Nearly all infectious λ particles formed in this way will contain an inserted piece of foreign DNA. Another advantage of using these modified viruses as vectors is that they enter bacteria much more easily than do plasmids. Among the variety of λ mutants that have been constructed for use as cloning vectors, one of them, called a *cosmid,* is essentially a hybrid of λ phage and a plasmid that can serve as a vector for large DNA inserts (as large as 45 kb).

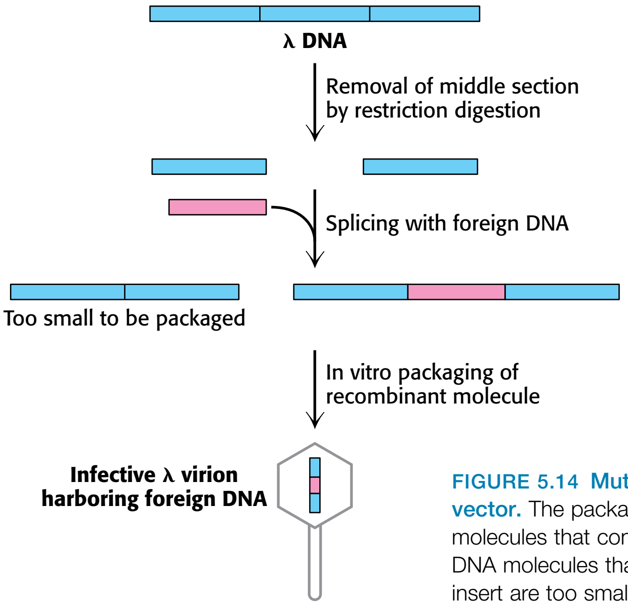

λ DNA
Removal of middle section by restriction digestion
Splicing with foreign DNA
Too small to be packaged
In vitro packaging of recombinant molecule
Infective λ virion harboring foreign DNA

FIGURE 5.14 Mutant λ phage as a cloning vector. The packaging process selects DNA molecules that contain an insert (colored red). DNA molecules that have resealed without an insert are too small to be efficiently packaged.

Bacterial and yeast artificial chromosomes

Much larger pieces of DNA can be propagated in *bacterial artificial chromosomes* (BACs) or *yeast artificial chromosomes* (YACs). BACs are highly engineered versions of the *E. coli* fertility (F) factor that can include inserts as large as 300 kb. YACs contain a centromere, an *autonomously replicating sequence* (ARS, where replication begins), a pair of telomeres (normal ends of eukaryotic chromosomes), selectable marker genes, and a cloning site (Figure 5.15). Inserts as large as 1000 kb can be cloned into YAC vectors.

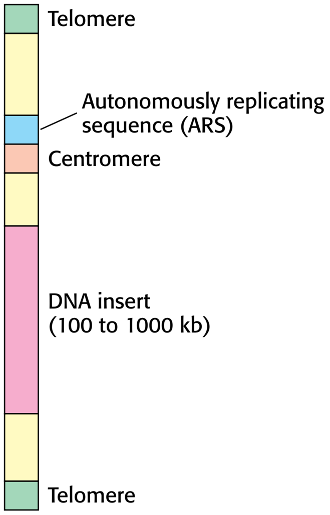

Telomere
Autonomously replicating sequence (ARS)
Centromere
DNA insert
(100 to 1000 kb)
Telomere

FIGURE 5.15 Diagram of a yeast artificial chromosome (YAC). These vectors include features necessary for replication and stability in yeast cells.

Specific genes can be cloned from digests of genomic DNA

Ingenious cloning and selection methods have made it possible to isolate small stretches of DNA in a genome containing more than 3×10^6 kb. The approach is to prepare a large collection (*library*) of DNA fragments and then to identify those members of the collection that have the gene of interest. Hence, to clone a gene that is present just once in an entire genome, two critical components must be available: a specific oligonucleotide probe for the gene of interest and a DNA library that can be screened rapidly.

How is a specific probe obtained? In one approach, a probe for a gene can be prepared if a part of the amino acid sequence of the protein encoded by the gene is known. Peptide sequencing of a purified protein (Chapter 3) or knowledge of the sequence of a homologous protein from a related species (Chapter 6) are two potential sources of such information. However, a problem arises because a single peptide sequence can be encoded by a number of different oligonucleotides (Figure 5.16). Thus, for this purpose, peptide sequences containing tryptophan and methionine are preferred, because these amino acids are specified by a single codon, whereas other amino acid residues have between two and six codons (Table 4.5). All the possible DNA sequences (or their complements) that encode the targeted peptide sequence are synthesized by the solid-phase method and made radioactive by phosphorylating their 5′ ends with ^{32}P.

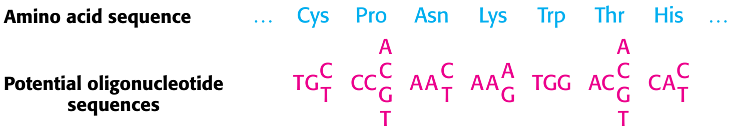

Amino acid sequence
… Cys Pro Asn Lys Trp Thr His …
Potential oligonucleotide sequences
TG C T
CC A C G T
AA C T
AA A G
TGG
AC A C G T
CA C T

FIGURE 5.16 Probes generated from a protein sequence. A probe can be generated by synthesizing all possible oligonucleotides encoding a particular sequence of amino acids. Because of the degeneracy of the genetic code, 256 distinct oligonucleotides must be synthesized to ensure that the probe matching the sequence of seven amino acids in this example is present.

To prepare the DNA library, a sample containing many copies of total genomic DNA is first mechanically sheared or partly digested by restriction enzymes

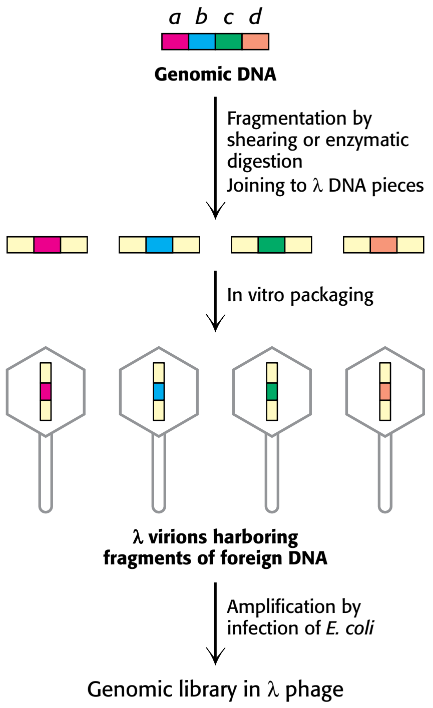

a *b* *c* *d*
Genomic DNA
Fragmentation by shearing or enzymatic digestion
Joining to λ DNA pieces
In vitro packaging
λ virions harboring fragments of foreign DNA
Amplification by infection of *E. coli*
Genomic library in λ phage

FIGURE 5.17 Creation of a genomic library. A genomic library can be created from a digest of a whole complex genome. After fragmentation of the genomic DNA into overlapping segments, the DNA is inserted into the λ phage vector (shown in yellow). Packaging into virions and amplification by infection in *E. coli* yields a genomic library.

into large fragments (Figure 5.17). This process yields a nearly random population of overlapping DNA fragments. These fragments are then separated by gel electrophoresis to isolate the set of all fragments that are about 15 kb long. Synthetic linkers are attached to the ends of these fragments, cohesive ends are formed, and the fragments are then inserted into a vector, such as λ phage DNA, prepared with the same cohesive ends. *E. coli* bacteria are then infected with these recombinant phages. These phages replicate themselves and then lyse their bacterial hosts. The resulting lysate contains fragments of human DNA housed in a sufficiently large number of virus particles to ensure that nearly the entire genome is represented. These phages constitute a *genomic library*. Phages can be propagated indefinitely such that the library can be used repeatedly over long periods.

This genomic library is then screened to find the very small number of phages harboring the gene of interest. For the human genome, a calculation shows that a 99% probability of success requires screening about 500,000 clones; hence, a very rapid and efficient screening process is essential. Rapid screening can be accomplished by DNA hybridization.

A dilute suspension of the recombinant phages is first plated on a lawn of bacteria (Figure 5.18). Where each phage particle has landed and infected a bacterium, a *plaque* containing identical phages develops on the plate. A replica of this master plate is then made by applying a sheet of nitrocellulose. Infected bacteria and phage DNA released from lysed cells adhere to the sheet in a pattern of spots corresponding to the plaques. Intact bacteria on this sheet are lysed with NaOH, which also serves to denature the DNA so that it becomes accessible for hybridization with a ^{32}P-labeled probe. The presence of a specific DNA sequence in a single spot on the replica can be detected by using a radioactive complementary DNA or RNA molecule as a probe. Autoradiography then reveals the positions of spots harboring recombinant DNA. The corresponding plaques are picked out of the intact master plate and grown. A single investigator can readily screen a million clones in a day. This method makes it possible to isolate virtually any gene, provided that a probe is available.

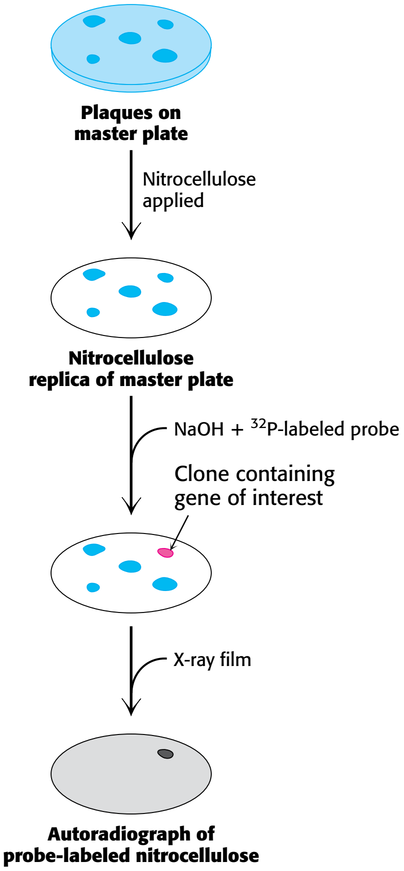

Plaques on master plate
Nitrocellulose applied
Nitrocellulose replica of master plate
NaOH + ^{32}P-labeled probe
Clone containing gene of interest
X-ray film
Autoradiograph of probe-labeled nitrocellulose

FIGURE 5.18 Screening a genomic library for a specific gene. Here, a plate is tested for plaques containing gene *a* of Figure 5.17.

Complementary DNA prepared from mRNA can be expressed in host cells

The preparation of eukaryotic DNA libraries presents unique challenges, especially if the researcher is interested primarily in the protein-coding region of a particular gene. Recall that most mammalian genes are mosaics of introns and exons. These interrupted genes cannot be expressed by bacteria, which lack the machinery to splice introns out of the primary transcript. However, this difficulty can be circumvented by causing bacteria to take up recombinant DNA that is complementary to mRNA, where the intronic sequences have been removed.

The key to forming *complementary DNA* is the enzyme *reverse transcriptase*. As discussed in Section 4.3, a retrovirus uses this enzyme to form a DNA–RNA hybrid in replicating its genomic RNA. Reverse transcriptase synthesizes a DNA strand complementary to an RNA template if the transcriptase is provided with a DNA primer that is base-paired to the RNA and contains a free 3′-OH group. We can use a simple sequence of linked thymidine [oligo(T)] residues as the primer. This oligo(T) sequence pairs with the poly(A) sequence at the 3′ end of most eukaryotic mRNA molecules (Section 4.4), as shown in Figure 5.19. The reverse transcriptase then synthesizes the rest of the cDNA strand in the presence of the four deoxyribonucleoside triphosphates (step 1). The RNA strand of this RNA–DNA hybrid is subsequently hydrolyzed by raising the pH (step 2). Unlike RNA, DNA is resistant to alkaline hydrolysis. The single-stranded DNA is converted into double-stranded DNA by creating another primer site. The enzyme *terminal transferase* adds nucleotides—for instance, several residues of dG—to the 3′ end of DNA (step 3). Oligo(dC) can bind to dG residues and prime the synthesis of the second DNA strand (step 4). Synthetic linkers can be added to this double-helical DNA for ligation to a suitable vector. Complementary DNA for all mRNA that a cell contains can be made, inserted into vectors, and then inserted into bacteria. Such a collection is called a *cDNA library*.

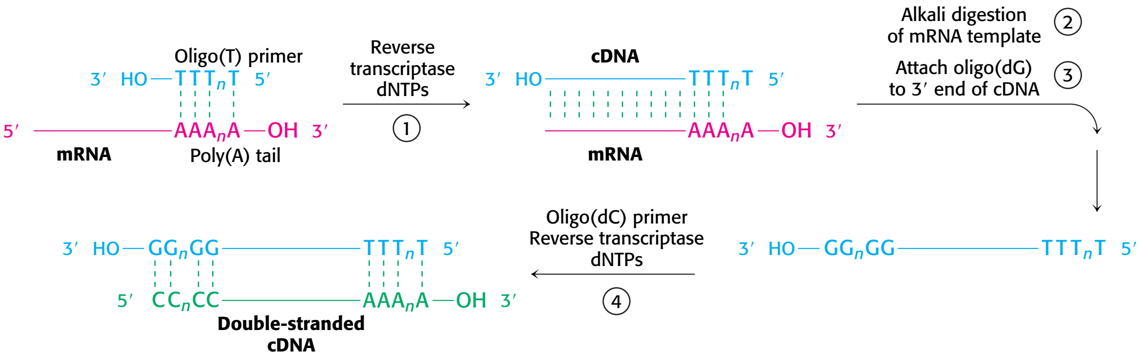

Oligo(T) primer
3′ HO—TTT$_n$T 5′
5′ —AAA$_n$A—OH 3′
mRNA
Poly(A) tail
Reverse transcriptase dNTPs
①
cDNA
3′ HO—TTT$_n$T 5′
—AAA$_n$A—OH 3′
mRNA
Alkali digestion of mRNA template ②
Attach oligo(dG) to 3′ end of cDNA ③
3′ HO—GG$_n$GG—TTT$_n$T 5′
Oligo(dC) primer
Reverse transcriptase
dNTPs
④
3′ HO—GG$_n$GG—TTT$_n$T 5′
5′ CC$_n$CC—AAA$_n$A—OH 3′
Double-stranded cDNA

FIGURE 5.19 Formation of a cDNA duplex. A complementary DNA (cDNA) duplex is created from mRNA by (1) use of reverse transcriptase to synthesize a cDNA strand, (2) digestion of the original RNA strand, (3) addition of several G bases to the DNA by terminal transferase, and (4) synthesis of a complementary DNA strand using the newly synthesized cDNA strand as a template.

Complementary DNA molecules can be inserted into expression vectors to enable the production of the corresponding protein of interest. Clones of cDNA can be screened on the basis of their capacity to direct the synthesis of a foreign protein in bacteria, a technique referred to as *expression cloning*. A labeled antibody specific for the protein of interest can be used to identify colonies of bacteria that express the corresponding protein product (Figure 5.20). As described earlier, spots of bacteria on a replica plate are lysed to release proteins, which bind to an applied nitrocellulose

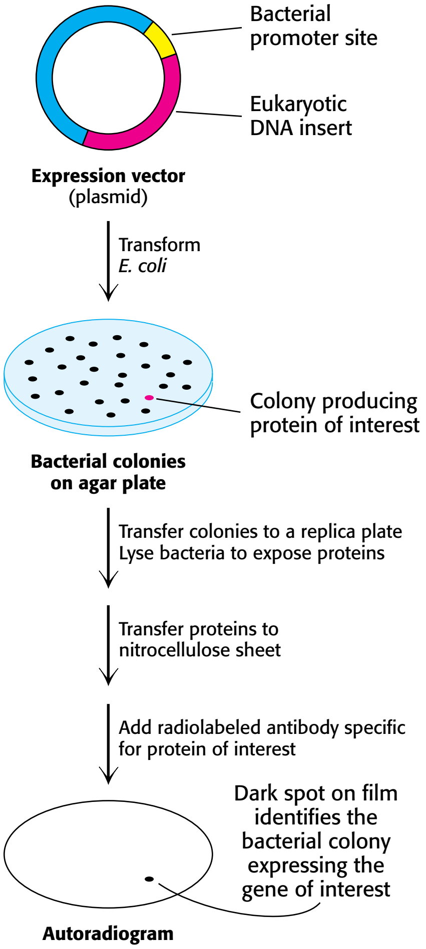

Bacterial promoter site
Eukaryotic DNA insert
Expression vector (plasmid)
Transform E. coli
Colony producing protein of interest
Bacterial colonies on agar plate
Transfer colonies to a replica plate
Lyse bacteria to expose proteins
Transfer proteins to nitrocellulose sheet
Add radiolabeled antibody specific for protein of interest
Dark spot on film identifies the bacterial colony expressing the gene of interest
Autoradiogram

FIGURE 5.20 Screening of cDNA clones. A method of screening for cDNA clones is to identify expressed products by staining with specific antibody.

filter. With the addition of labeled antibody specific for the protein of interest, the location of the desired colonies on the master plate can be readily identified. This immunochemical screening approach can be used whenever a protein is expressed and corresponding antibody is available.

Complementary DNA has many applications beyond the generation of genetic libraries. The overproduction and purification of most eukaryotic proteins in prokaryotic cells necessitates the insertion of cDNA into plasmid vectors. For example, proinsulin, a precursor of insulin, is synthesized by bacteria-harboring plasmids that contain DNA complementary to mRNA for proinsulin (Figure 5.21). Indeed, bacteria produce much of the insulin used today by millions of diabetics.

Proteins with new functions can be created through directed changes in DNA

Much has been learned about genes and proteins by analyzing the effects that mutations have on their structure and function. In the classic genetic approach, mutations are generated randomly throughout the genome of a host organism, and those individuals exhibiting a phenotype of interest are selected. Analysis of these mutants then reveals which genes are altered, and DNA sequencing identifies the precise nature of the changes. *Recombinant DNA technology makes the creation of specific mutations feasible in vitro.* We can construct new genes with designed properties by making three kinds of directed changes: *deletions*, *insertions*, and *substitutions*. A variety of methods can be used to introduce these types of mutations, including the following examples.

Site-directed mutagenesis. Mutant proteins with single amino acid substitutions can be readily produced by *site-directed mutagenesis* (Figure 5.22). Suppose that we want to replace a particular serine residue with cysteine. This mutation can be made if (1) we have a plasmid containing the gene or cDNA for the protein and (2) we know the base sequence around the site to be altered. If the serine of interest is encoded by TCT, mutation of the central base from C to G yields the TGT codon, which encodes cysteine. This type of mutation is called a *point mutation* because only one base is altered. To introduce this mutation into our plasmid, we prepare an oligonucleotide primer that is complementary to this region of the gene except that it contains TGT instead of TCT. The two strands of the plasmid are separated, and the primer is then annealed to the complementary strand. The mismatch of 1 of 15 base pairs is tolerable if the annealing is carried out at an appropriate temperature. After annealing to the complementary strand, the primer is elongated by DNA polymerase, and the double-stranded circle is closed by adding DNA ligase. Subsequent replication of this duplex yields two kinds of progeny plasmid, half with the original TCT sequence and half with the mutant TGT sequence. Expression of the plasmid containing the new TGT

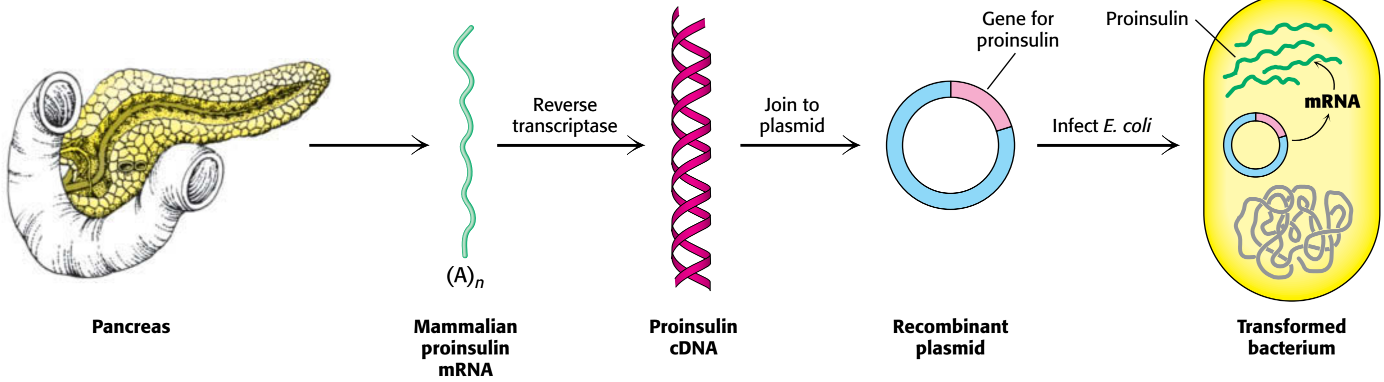

Pancreas
Reverse transcriptase
(A)$_n$
Mammalian proinsulin mRNA
Join to plasmid
Proinsulin cDNA
Gene for proinsulin
Recombinant plasmid
Infect E. coli
Proinsulin
mRNA
Transformed bacterium

FIGURE 5.21 Synthesis of proinsulin by bacteria. Proinsulin, a precursor of insulin, can be synthesized by transformed (genetically altered) clones of *E. coli*. The clones contain the mammalian proinsulin gene.

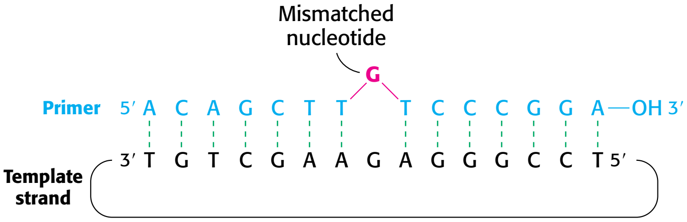

Mismatched nucleotide
G
Primer
5′ A C A G C T T T C C C G G A—OH 3′
3′ T G T C G A A G A G G G C C T 5′
Template strand

FIGURE 5.22 Oligonucleotide-directed mutagenesis. A primer containing a mismatched nucleotide is used to produce a desired change in the DNA sequence.

sequence will produce a protein with the desired substitution of cysteine for serine at a unique site. We will encounter many examples of the use of site-directed mutagenesis to precisely alter regulatory regions of genes and to produce proteins with tailor-made features.

Cassette mutagenesis. In *cassette mutagenesis,* a variety of mutations, including insertions, deletions, and multiple point mutations, can be introduced into the gene of interest. A plasmid harboring the original gene is cut with a pair of restriction enzymes to remove a short segment (Figure 5.23). A synthetic double-stranded oligonucleotide—the *cassette*—carrying the genetic alterations of interest is prepared with cohesive ends that are complementary to the ends of the cut plasmid. Ligation of the cassette into the plasmid yields the desired mutated gene product.

Mutagenesis by PCR. In Section 5.1, we learned how PCR can be used to amplify a specific region of DNA using primers that flank the region of interest. In fact, the creative design of PCR primers enables the introduction of specific insertions, deletions, and substitutions into the amplified sequence. A number of methods have been developed for this purpose. Here, we shall consider one: *inverse PCR* to introduce deletions into plasmid DNA (Figure 5.24). In this approach, primers are designed to flank the sequence to be deleted. However, these primers are oriented in the opposite

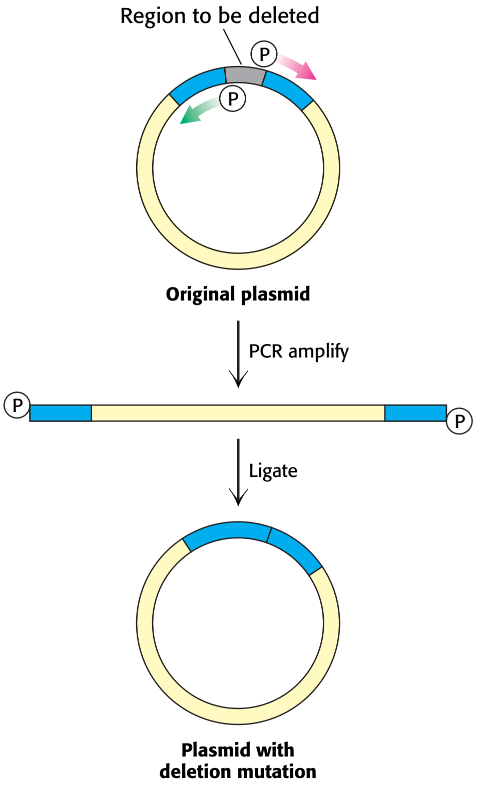

Region to be deleted
P
P
Original plasmid
PCR amplify
P
P
Ligate
Plasmid with deletion mutation

FIGURE 5.24 Deletion mutagenesis by inverse PCR. A deletion can be introduced into a plasmid with primers that flank this region but are oriented away from the segment to be removed. PCR amplification yields a linear product that contains the entire plasmid minus the unwanted sequence. If the primers contained a 5' phosphate group, this product can be recircularized using DNA ligase, generating a plasmid with the desired mutation.

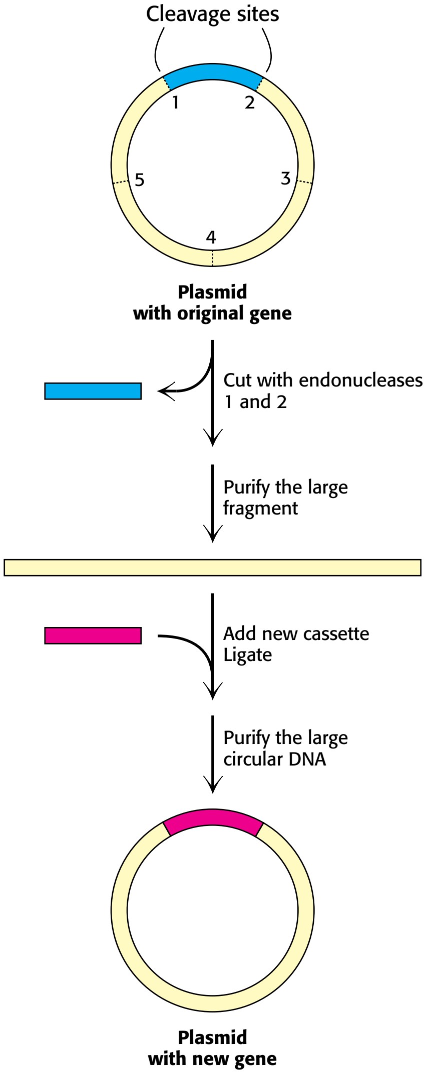

Cleavage sites
1
2
3
4
5
Plasmid with original gene
Cut with endonucleases 1 and 2
Purify the large fragment
Add new cassette
Ligate
Purify the large circular DNA
Plasmid with new gene

FIGURE 5.23 Cassette mutagenesis. DNA is cleaved at a pair of unique restriction sites by two different restriction endonucleases. A synthetic oligonucleotide with ends that are complementary to these sites (the cassette) is then ligated to the cleaved DNA. The method is highly versatile because the inserted DNA can have any desired sequence.

direction, such that they direct the amplification of the entire plasmid, minus the region to be deleted. If each of the primers contains a 5′phosphate group, the amplified product can be recircularized with DNA ligase, yielding the desired deletion mutation.

Designer genes. Novel proteins can also be created by splicing together gene segments that encode domains that are not associated in nature. For example, a gene for an antibody can be joined to a gene encoding a toxic protein, yielding a chimeric protein that kills those cells recognized by the antibody. These *immunotoxins* are being evaluated as anticancer agents. Furthermore, noninfectious coat proteins of viruses can be produced in large amounts by recombinant DNA methods. They can serve as *synthetic vaccines* that are safer than conventional vaccines prepared by inactivating pathogenic viruses. A subunit of the hepatitis B virus produced in yeast is proving to be an effective vaccine against this debilitating viral disease. Finally, entirely new genes can be synthesized de novo by the solid-phase method described above. These genes can encode proteins with no known counterparts in nature.

Recombinant methods enable the exploration of the functional effects of disease-causing mutations

The application of recombinant DNA technology to the production of mutated proteins has had a significant effect in the study of ALS. Recall that genetic studies had identified a number of ALS-inducing mutations within the gene encoding Cu/Zn superoxide dismutase. As we shall learn in Section 18.3, SOD1 catalyzes the conversion of the superoxide radical anion into hydrogen peroxide, which, in turn, is converted into molecular oxygen and water by catalase. To study the potential effect of ALS-causing mutations on SOD1 structure and function, the *SOD1* gene was isolated from a human cDNA library by PCR amplification. The amplified fragments containing the gene were then digested by an appropriate restriction enzyme and inserted into a similarly digested plasmid vector. Mutations corresponding to those observed in ALS patients were introduced into these plasmids by oligonucleotide-directed mutagenesis and the protein products were expressed and assayed for their catalytic activity. Surprisingly, these mutations did not significantly alter the enzymatic activity of the corresponding recombinant proteins. These observations have led to the prevailing notion that these mutations impart toxic properties to SOD1. Although the nature of this toxicity is not yet completely understood, one hypothesis is that mutant SOD1 is prone to form toxic aggregates in the cytoplasm of neuronal cells.

5.3 Complete Genomes Have Been Sequenced and Analyzed

The methods just described are extremely effective for the isolation and characterization of fragments of DNA. However, the genomes of organisms ranging from viruses to human beings contain considerably longer sequences, arranged in very specific ways crucial for their integrated functions. Is it possible to sequence complete genomes and analyze them? For small genomes, this sequencing was accomplished soon after DNA-sequencing methods were developed. Sanger and his coworkers determined the complete sequence of the 5386 bases in the genome of the φX174 DNA virus in 1977, just a quarter century after Sanger's pioneering elucidation of the amino acid sequence of a protein. This tour de force was followed several years later by the determination of the sequence of human mitochondrial DNA, a double-stranded circular DNA molecule containing 16,569 base pairs. It encodes 2 ribosomal RNAs, 22 transfer RNAs, and 13 proteins. Many other viral genomes were sequenced in subsequent years. However, the genomes of free-living organisms presented a great challenge because even the simplest comprises more than 1 million base pairs. Thus, sequencing projects require both rapid sequencing techniques and efficient methods for assembling many short stretches of 300 to 500 base pairs into a complete sequence.

The genomes of organisms ranging from bacteria to multicellular eukaryotes have been sequenced

With the development of automatic DNA sequencers based on fluorescent dideoxynucleotide chain terminators, high-volume, rapid DNA sequencing became a reality. The genome sequence of the bacterium *Haemophilus influenzae* was determined in 1995 by using a "shotgun" approach. The genomic DNA was sheared randomly into fragments that were then sequenced. Computer programs assembled the complete sequence by matching up overlapping regions between fragments. The *H. influenzae* genome comprises 1,830,137 base pairs and encodes approximately 1740 proteins (Figure 5.25). Using similar approaches, as well as more advanced methods described below, investigators have determined the sequences of more than 10,000 bacterial and archaeal species, including key model organisms such as *E. coli, Salmonella typhimurium,* and *Archaeoglobus fulgidus,* as well as pathogenic organisms such as *Yersina pestis* (bubonic plague) and *Bacillus anthracis* (anthrax).

The first eukaryotic genome to be completely sequenced was that of baker's yeast, *Saccharomyces cerevisiae,* in 1996. The yeast genome comprises approximately 12 million base pairs, distributed on 16 chromosomes, and encodes more than 6000 proteins. This achievement was followed in 1998 by the first complete sequencing of the genome of a multicellular organism, the nematode *Caenorhabditis elegans,* which contains 97 million base pairs. This genome includes more than 19,000 genes. The genomes of

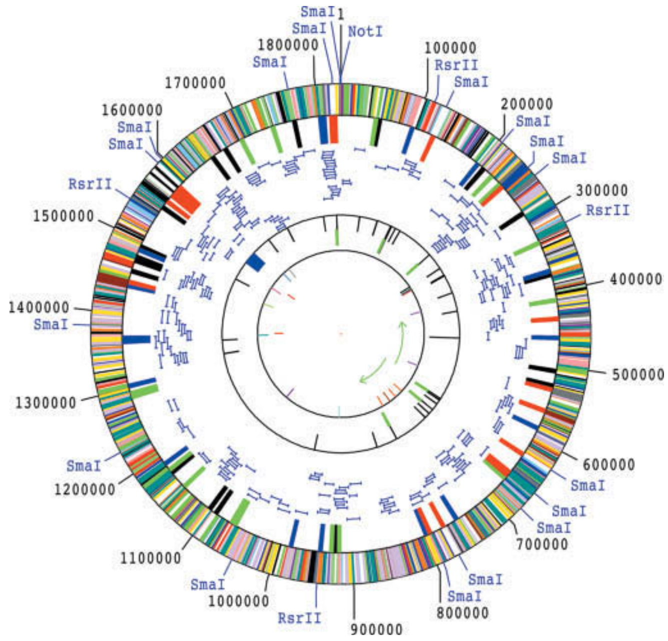


FIGURE 5.25 A complete genome. The diagram depicts the genome of *Haemophilus influenzae,* the first complete genome of a free-living organism to be sequenced. The genome encodes more than 1700 proteins and 70 RNA molecules. The likely function of approximately one-half of the proteins was determined by comparisons with sequences of proteins already characterized in other species. [Data from R. D. Fleischmann et al., *Science* 269:496–512, 1995; scan courtesy of The Institute for Genomic Research.]

many additional organisms widely used in biological and biomedical research have now been sequenced, including those of the fruit fly *Drosophila melanogaster,* the model plant *Arabidopsis thaliana,* the mouse, the rat, and the dog. Note that even after a genome sequence has been considered complete, some sections, such as the repetitive sequences that make up heterochromatin, may be missing because these DNA sequences are very difficult to manipulate with the use of standard techniques.

The sequence of the human genome has been completed

The ultimate goal of much of genomics research has been the sequencing and analysis of the human genome. Given that the human genome comprises approximately 3 billion base pairs of DNA distributed among 24 chromosomes, the challenge of producing a complete sequence was daunting. However, through an organized international effort of academic laboratories and private companies, the human genome has progressed from a draft sequence first reported in 2001 to a finished sequence reported in late 2004 (Figure 5.26).

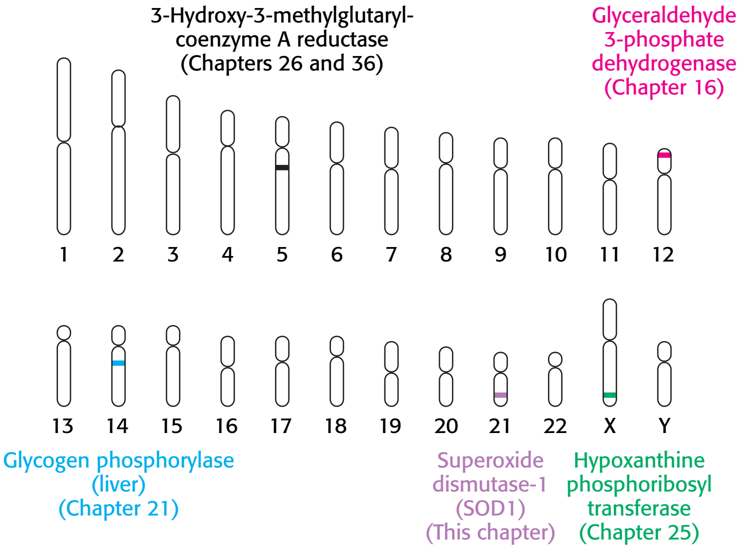

3-Hydroxy-3-methylglutaryl-coenzyme A reductase (Chapters 26 and 36)
Glyceraldehyde 3-phosphate dehydrogenase (Chapter 16)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Glycogen phosphorylase (liver) (Chapter 21)
Superoxide dismutase-1 (SOD1) (This chapter)
Hypoxanthine phosphoribosyl transferase (Chapter 25)

FIGURE 5.26 The human genome. The human genome is arrayed on 46 chromosomes—22 pairs of autosomes and the X and Y sex chromosomes. The locations of several genes associated with important pathways in biochemistry are highlighted.

The human genome is a rich source of information about many aspects of humanity, including biochemistry and evolution. Analysis of the genome will continue for many years to come. Developing an inventory of protein-encoding genes is one of the first tasks. At the beginning of the genome-sequencing project, the number of such genes was estimated to be approximately 100,000. With the availability of the completed (but not finished) genome, this estimate was reduced to between 30,000 and 35,000. With the finished sequence, the estimate fell to between 20,000 to 25,000. We will use the estimate of 23,000 throughout this book. The reduction in this estimate is due, in part, to the realization that there are a large number of *pseudogenes,* formerly functional genes that have accumulated mutations such that they no longer produce proteins. For example, more than half of the genomic regions that correspond to olfactory receptors—key molecules responsible for our sense of smell—are pseudogenes (Section 33.1). The corresponding regions in the genomes of other primates and rodents encode functional olfactory receptors. Nonetheless, the surprisingly small number of genes belies the complexity of the human proteome. *Many genes encode more than one protein through mechanisms such as alternative splicing of mRNA and posttranslational modifications of proteins.* The different proteins encoded by a single gene often display important variations in functional properties.

The human genome contains a large amount of DNA that does not encode proteins. A great challenge in modern biochemistry and genetics is to elucidate the roles of this noncoding DNA. Much of this DNA is present because of the existence of *mobile genetic elements.* These elements, related to retroviruses (Section 4.3), have inserted themselves throughout the genome over time. Most of these elements have accumulated mutations and are no longer functional. For example, more than 1 million *Alu sequences,* each approximately 300 bases in length, are present in the human genome. *Alu* sequences are examples of *SINES, short interspersed elements.* The human genome also includes nearly 1 million *LINES, long interspersed elements,* DNA sequences that can be as long as 10 kilobases (kb). The roles of these elements as neutral genetic parasites or instruments of genome evolution are under current investigation.

Next-generation sequencing methods enable the rapid determination of a complete genome sequence

Since the introduction of Sanger dideoxy method in the mid-1970s, significant advances have been made in DNA-sequencing technologies, enabling the readout of progressively longer sequences with higher fidelity and shorter run times. The development of *next-generation sequencing* (NGS) platforms has extended this capability to formerly unforeseen levels. By combining technological breakthroughs in the handling of very small amounts of liquid, high-resolution optics, and computing power, these methods have already made a significant impact on the ability to obtain whole genome sequences rapidly and cheaply (Chapter 1).

Next-generation sequencing refers to a family of technologies, each of which utilizes a unique approach for the determination of a DNA sequence. All of these methods are *highly parallel*: from 1 million to 1 billion DNA fragment sequences are acquired in a single experiment. How are NGS methods capable of attaining such a high number of parallel runs? Individual DNA fragments are amplified by PCR on a solid support—a single bead or a small region of a glass slide—such that clusters of identical DNA fragments are distinguishable by high-resolution imaging. These fragments then serve as templates for DNA polymerase, where the addition of nucleotide triphosphates is converted to a signal that can be detected in a highly sensitive manner. The technique used to detect individual base incorporation varies among the variety of NGS methods. However, most of these can be understood simply by considering the overall reaction of chain elongation catalyzed by DNA polymerase (Figure 5.27). In the *reversible terminator method*, the four nucleotides are added to the template DNA, with each base tagged with a unique fluorescent label and a reversibly-blocked 3′ end. The blocked end assures that only one phosphodiester linkage will form. Once the nucleotide is incorporated into the growing strand, it is identified by its fluorescent tag, the blocking agent is removed, and the process is repeated.

In *pyrosequencing*, nucleotides are added to the template DNA, one at a time in a defined order. One of the nucleotides will be incorporated into the growing strand, releasing a pyrophosphate which is detected by coupling

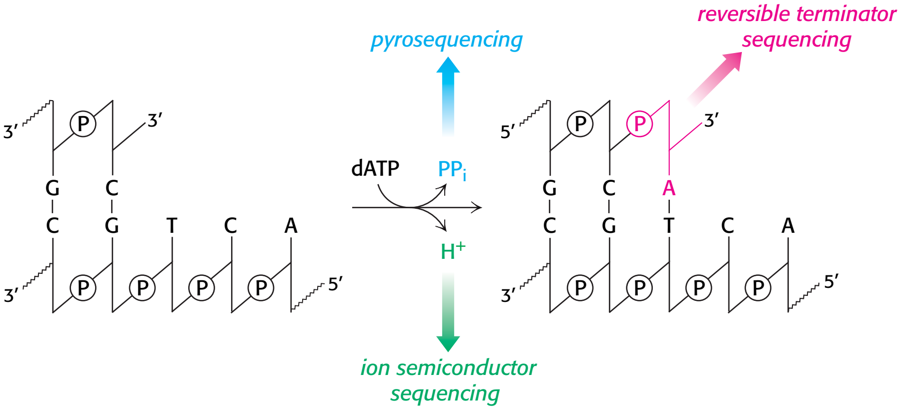


FIGURE 5.27 Detection methods in next-generation sequencing. Measurement of base incorporation in next-generation sequencing methods relies on the detection of the various products of the DNA polymerase reaction. Reversible terminator sequencing measures the nucleotide incorporation in a manner similar to Sanger sequencing, while pyrosequencing and ion semiconductor sequencing detect the release of pyrophosphate and protons, respectively.

the formation of pyrophosphate with the production of light by the sequential action of the enzymes ATP sulfurylase and luciferase:

$$\mathrm{PP_i} + \text{adenylyl sulfate} \xrightleftharpoons{\text{ATP sulfurylase}} \text{ATP} + \text{sulfate}$$

$$\text{ATP} + \text{luciferin} \xrightleftharpoons{\text{luciferase}} \text{oxyluciferin} + \textbf{light}$$

The protocol for *ion semiconductor sequencing* is similar to pyrosequencing except that nucleotide incorporation is detected by sensitively measuring the very small changes in pH of the reaction mixture due to the release of proton upon nucleotide incorporation.

Regardless of the sequencing method, the technology exists to quantify the signal produced by millions of DNA fragment templates simultaneously. However, for many approaches, as few as 50 bases are read per fragment. Hence, significant computing power is required to both store the massive amounts of sequence data and perform the necessary alignments required to assemble a completed sequence. NGS methods are being used to answer an ever-growing number of questions in genomics, transcriptomics, and evolutionary biology, to name a few. Additionally, individual genome sequences will provide information about genetic variation within populations and may usher in an era of personalized medicine, when these data can be used to guide treatment decisions.

Comparative genomics has become a powerful research tool

Comparisons with genomes from other organisms are sources of insight into the human genome. The sequencing of the genome of the chimpanzee, our closest living relative, as well as that of other mammals that are widely used in biological research, such as the mouse and the rat, have been completed. Comparisons reveal that an astonishing 99% of human genes have counterparts in these rodent genomes. However, these genes have been substantially reassorted among chromosomes in the estimated 75 million years of evolution since humans and rodents had a common ancestor (Figure 5.28).

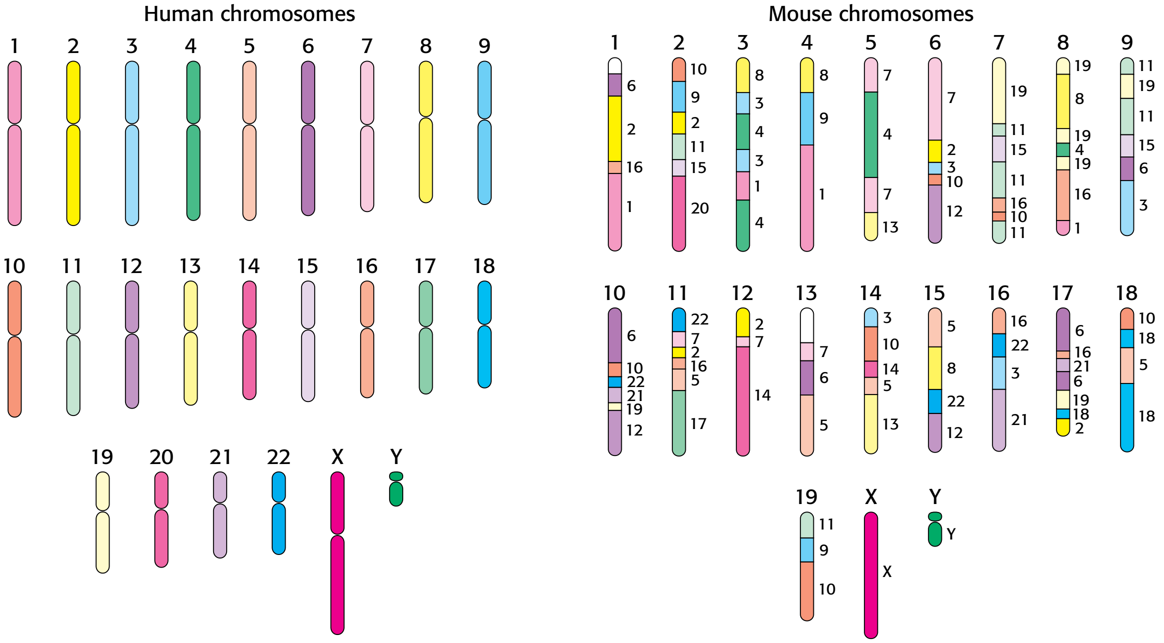

Human chromosomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Mouse chromosomes
1: 6, 2, 16, 1
2: 10, 9, 2, 11, 15, 20
3: 8, 3, 4, 3, 1, 4
4: 8, 9, 1
5: 7, 4, 7, 13
6: 7, 2, 3, 10, 12
7: 19, 11, 15, 11, 16, 10, 11
8: 19, 8, 19, 4, 19, 16, 1
9: 11, 19, 11, 15, 6, 3
10: 6, 10, 22, 21, 19, 12
11: 22, 7, 2, 16, 5, 17
12: 2, 7, 14
13: 7, 6, 5
14: 3, 10, 14, 5, 13
15: 5, 8, 22, 12
16: 16, 22, 3, 21
17: 6, 16, 21, 6, 19, 18, 2
18: 10, 18, 5, 18
19: 11, 9, 10
X: X
Y: Y

FIGURE 5.28 Genome comparison. A schematic comparison of the human genome and the mouse genome shows reassortment of large chromosomal fragments. The small numbers to the right of the mouse chromosomes indicate the human chromosome to which each region is most closely related.

The genomes of other organisms also have been determined specifically for use in comparative genomics. For example, the genomes of two species of puffer fish, *Takifugu rubripes* and *Tetraodon nigroviridis,* have been determined. These genomes were selected because they are very small and lack much of the intergenic DNA present in such abundance in the human genome. The puffer fish genomes include fewer than 400 megabase pairs (Mbp), one-eighth of the number in the human genome, yet the puffer fish and human genomes contain essentially the same number of genes. Comparison of the genomes of these species with that of humans revealed more than 1000 formerly unrecognized human genes. Furthermore, comparison of the two species of puffer fish, which had a common ancestor approximately 25 million years ago, is a source of insight into more-recent events in evolution. Comparative genomics is a powerful tool, both for interpreting the human genome and for understanding major events in the origin of genera and species.

A puffer fish. [Beth Swanson/Shutterstock]

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

After a gene of interest has been identified, cloned, and sequenced, it is often desirable to understand how that gene and its corresponding protein product function in the context of a whole cell or organism. It is now possible to determine how the expression of a particular gene is regulated, how mutations in the gene affect the function of the corresponding protein product, and how the behavior of an entire cell or model organism is altered by the introduction of mutations within specific genes. Levels of transcription of large families of genes within cells and tissues can be readily quantitated and compared across a range of environmental conditions. Eukaryotic genes can be introduced into bacteria, and the bacteria can be used as factories to produce a desired protein product. DNA can also be introduced into the cells of higher organisms. Genes introduced into animals are valuable tools for examining gene action, and they are the basis of gene therapy. Genes introduced into plants can make the plants resistant to pests, able to grow in harsh conditions, or carry greater quantities of essential nutrients. The manipulation of eukaryotic genes holds much promise as a source of medical and agricultural benefits.

Gene-expression levels can be comprehensively examined

Most genes are present in the same quantity in every cell—namely, one copy per haploid cell or two copies per diploid cell. However, the level at which a gene is expressed, as indicated by mRNA quantities, can vary widely, ranging from no expression to hundreds of mRNA copies per cell. Gene-expression patterns vary from cell type to cell type, distinguishing, for example, a muscle cell from a nerve cell. Even within the same cell, gene-expression levels may vary as the cell responds to changes in physiological circumstances. Note that mRNA levels sometimes correlate with the levels of proteins expressed, but this correlation does not always hold. Thus, care must be exercised when interpreting the results of mRNA levels alone.

The quantity of individual mRNA transcripts can be determined by *quantitative PCR* (qPCR), or real-time PCR. RNA is first isolated from the cell or tissue of interest. With the use of reverse transcriptase, cDNA is prepared from this RNA sample. In one qPCR approach, the transcript of interest is PCR amplified with the appropriate primers in the presence of the dye SYBR Green I, which fluoresces brightly when bound to

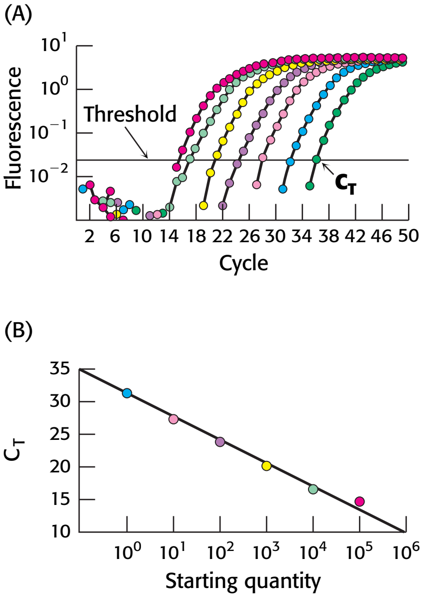


FIGURE 5.29 Quantitative PCR. (A) In qPCR, fluorescence is monitored in the course of PCR amplification to determine C_T, the cycle at which this signal exceeds a defined threshold. Each color represents a different starting quantity of DNA. (B) C_T values are inversely proportional to the number of copies of the original cDNA template. [Data from N. J. Walker, *Science* 296:557–559, 2002.]

double-stranded DNA. In the initial PCR cycles, not enough duplex is present to allow a detectable fluorescence signal. However, after repeated PCR cycles, the fluorescence intensity exceeds the detection threshold and continues to rise as the number of duplexes corresponding to the transcript of interest increases (Figure 5.29). Importantly, the cycle number at which the fluorescence becomes detectable over a defined threshold (or C_T) is indirectly proportional to the number of copies of the original template. After the relation between the original copy number and the C_T has been established with the use of a known standard, subsequent qPCR experiments can be used to determine the number of copies of any desired transcript in the original sample, provided the appropriate primers are available.

Although qPCR is a powerful technique for quantitation of a small number of transcripts in any given experiment, we can now use our knowledge of complete genome sequences to investigate an entire *transcriptome,* the pattern and level of expression of all genes in a particular cell or tissue. One of the most powerful methods for this purpose is based on hybridization. Single-stranded oligonucleotides whose sequences correspond to coding regions of the genome are affixed to a solid support such as a microscope slide, creating a *DNA microarray*. Importantly, the position of each sequence within the array is known. mRNA is isolated from the cells of interest (a tumor, for example) as well as a control sample (Figure 5.30). From this mRNA, cDNA is prepared (Section 5.2) in the presence of fluorescent nucleotides using different labels, usually green and red, for the two samples. The samples are combined, separated into single strands, and hybridized to the slide. The relative levels of green and red fluorescence at each

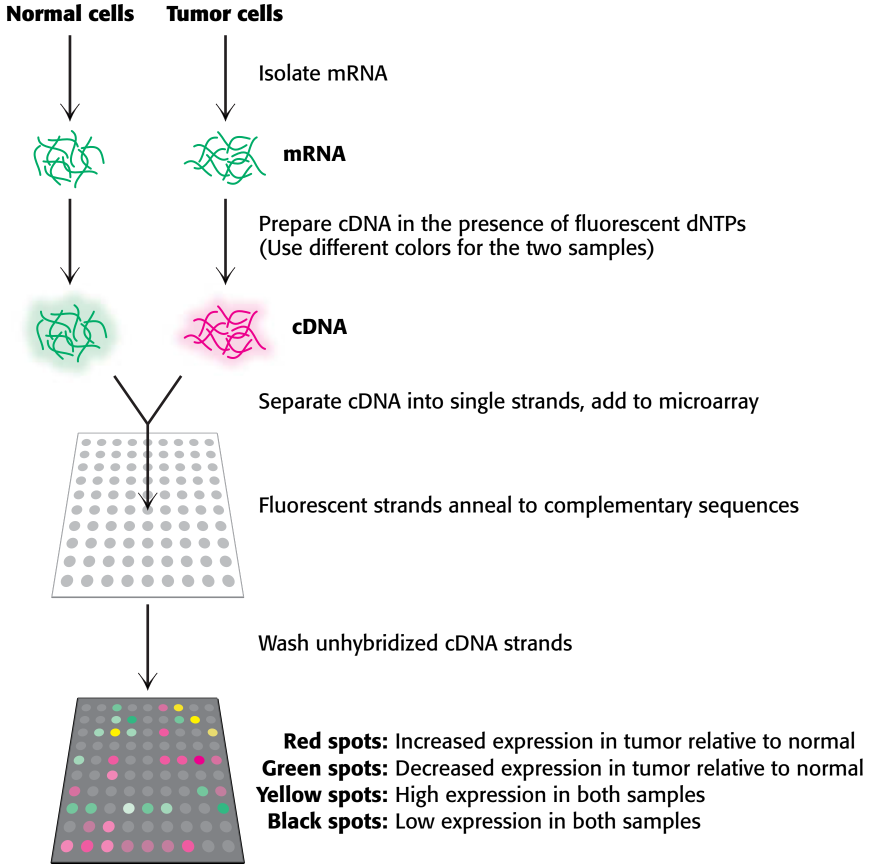

Normal cells
Tumor cells
Isolate mRNA
mRNA
Prepare cDNA in the presence of fluorescent dNTPs
(Use different colors for the two samples)
cDNA
Separate cDNA into single strands, add to microarray
Fluorescent strands anneal to complementary sequences
Wash unhybridized cDNA strands
Red spots: Increased expression in tumor relative to normal
Green spots: Decreased expression in tumor relative to normal
Yellow spots: High expression in both samples
Black spots: Low expression in both samples

FIGURE 5.30 Using DNA microarrays to measure gene expression changes in a tumor. mRNA is isolated from two samples, tumor cells and a control sample. From these transcripts, cDNA is prepared in the presence of a fluorescent nucleotide, with a red label for the tumor sample and a green label for the control sample. The cDNA strands are separated, hybridized to the microarray, and the unbound DNA is washed away. Spots that are red indicate genes which are expressed more highly in the tumor, while the green spots indicate reduced expression relative to control. Spots that are black or yellow indicate comparable expression at either low or high levels, respectively. [Information from D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry* 6th ed. (W. H. Freeman and Company, 2013)]

spot indicate the differences in expression for each gene. DNA chips have been prepared such that thousands of transcript levels can be assessed in a single experiment. Hence, over several arrays, the differences in expression of many genes across a number of different cell types or conditions can be measured (Figure 5.31).

Microarray analyses can be quite informative in the study of gene-expression changes in diseased mammals compared with their healthy counterparts. As noted earlier, although ALS-causing mutations within the *SOD1* gene had been identified, the mechanism by which the mutant SOD1 protein ultimately leads to motor-neuron loss remains a mystery. Many research groups have used microarray analysis of neuronal cells isolated from humans and mice carrying *SOD1* gene mutations to search for clues into the pathways of disease progression and to suggest potential avenues for treatment. These studies have implicated a variety of biochemical pathways, including immunological activation, handling of oxidative stress, and protein degradation, in the cellular response to the mutant, toxic forms of SOD1.

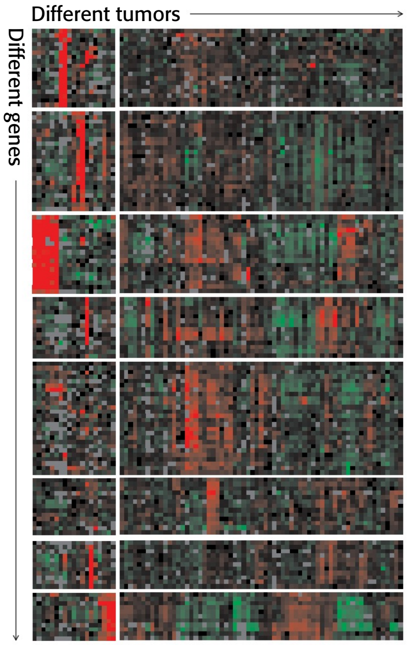


FIGURE 5.31 Gene-expression analysis with the use of microarrays. The expression levels of thousands of genes can be simultaneously analyzed with DNA microarrays. Here, an analysis of 1733 genes in 84 breast tumor samples reveals that the tumors can be divided into distinct classes based on their gene-expression patterns. In this "heat map" representation, each row represents a different gene and each column represents a different breast tumor sample (i.e., a separate microarray experiment). Red corresponds to gene induction and green corresponds to gene repression. [Data from C. M. Perou et al., *Nature* 406:747–752, 2000.]

New genes inserted into eukaryotic cells can be efficiently expressed

Bacteria are ideal hosts for the amplification of DNA molecules. They can also serve as factories for the production of a wide range of prokaryotic and eukaryotic proteins. However, bacteria lack the necessary enzymes to carry out posttranslational modifications such as the specific cleavage of polypeptides and the attachment of carbohydrate units. Thus, many eukaryotic genes can be expressed correctly only in eukaryotic host cells. The introduction of recombinant DNA molecules into cells of higher organisms can also be a source of insight into how their genes are organized and expressed. How are genes turned on and off in embryological development? How does a fertilized egg give rise to an organism with highly differentiated cells that are organized in space and time? These central questions of biology can now be fruitfully approached by expressing foreign genes in mammalian cells.

Recombinant DNA molecules can be introduced into animal cells in several ways. In one method, foreign DNA molecules precipitated by calcium phosphate are taken up by animal cells. A small fraction of the imported DNA becomes stably integrated into the chromosomal DNA. The efficiency of incorporation is low, but the method is useful because it is easy to apply. In another method, DNA is *microinjected* into cells. A fine-tipped glass micropipette containing a solution of foreign DNA is inserted into a nucleus (Figure 5.32). A skilled investigator can inject hundreds of cells per hour. About 2% of injected mouse cells are viable and contain the new gene. In a third method, *viruses* are used to introduce new genes into animal cells. The most effective vectors are *retroviruses,* whose genomes are encoded by RNA and replicate through DNA intermediates. A striking feature of the life cycle of a retrovirus is that the double-helical DNA form of its genome, produced by the action of reverse transcriptase, becomes randomly incorporated into host chromosomal DNA. This DNA version of the viral genome, called *proviral DNA,* can be efficiently expressed by the host cell and replicated along with normal cellular DNA. Retroviruses do not usually kill their hosts. Foreign genes have been efficiently introduced into mammalian cells by infecting them with vectors derived from the *Moloney murine leukemia virus,* a retrovirus which can accept inserts as long as 6 kb. Some genes introduced by this vector into the genome of a transformed host cell are efficiently expressed.

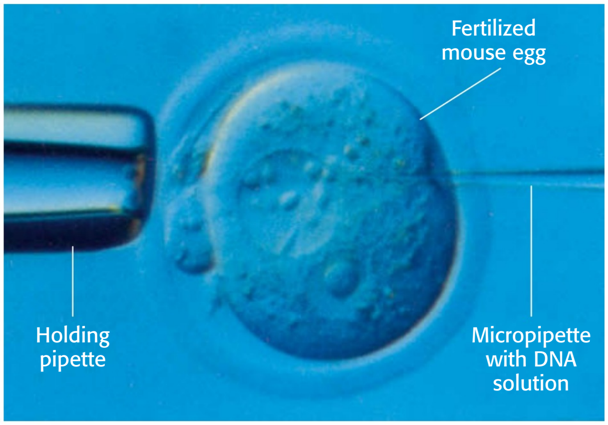


FIGURE 5.32 Microinjection of DNA. Cloned plasmid DNA is being microinjected into the male pronucleus of a fertilized mouse egg.

Two other viral vectors are extensively used. *Vaccinia virus,* a large DNA-containing virus, replicates in the cytoplasm of mammalian cells, where it shuts down host-cell protein synthesis. *Baculovirus* infects insect cells, which can be conveniently cultured. Insect larvae infected with this virus can serve as efficient protein factories. Vectors based on these large-genome viruses have been engineered to express DNA inserts efficiently.

Transgenic animals harbor and express genes introduced into their germ lines

As shown in Figure 5.32, plasmids harboring foreign genes can be microinjected into the male pronucleus of fertilized mouse eggs, which are then inserted into the uterus of a foster-mother mouse. A subset of the resulting embryos in this host will then harbor the foreign gene; these embryos may develop into mature animals. Southern blotting or PCR analysis of DNA isolated from the progeny can be used to determine which offspring carry the introduced gene. These *transgenic mice* are a powerful means of exploring the role of a specific gene in the development, growth, and behavior of an entire organism. Transgenic animals often serve as useful models for a particular disease process, enabling researchers to test the efficacy and safety of a newly developed therapy.

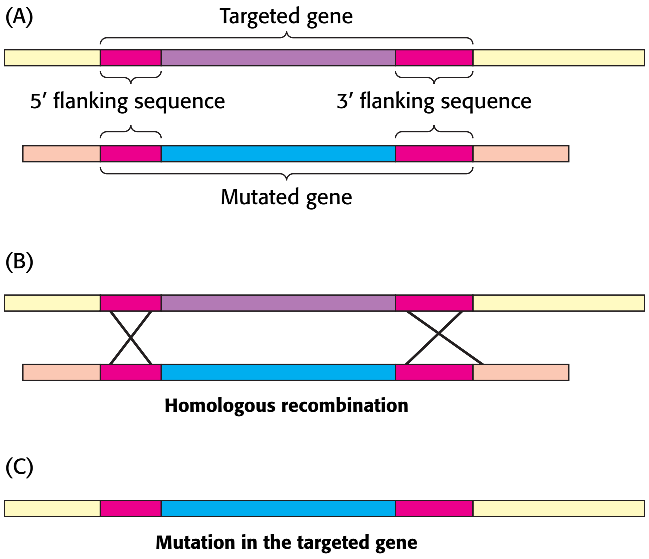
Let us return to our example of ALS. Research groups have generated transgenic mouse lines that express forms of human superoxide dismutase that harbor mutations matching those identified in earlier genetic analyses. Many of these strains exhibit a clinical picture similar to that observed in ALS patients: progressive weakness of voluntary muscles and eventual paralysis, motor-neuron loss, and rapid progression to death. Since their first characterization in 1994, these strains continue to serve as valuable sources of information for the exploration of the mechanism, and potential treatment, of ALS.

Gene disruption and genome editing provide clues to gene function and opportunities for new therapies

The function of a gene can also be probed by inactivating it and looking for resulting abnormalities. Powerful methods have been developed for accomplishing *gene disruption* (also called *gene knockout*) in organisms such as yeast and mice. These methods rely on the process of *homologous recombination* (Section 28.5), in which two DNA molecules with strong sequence similarity exchange segments. If a region of foreign DNA is flanked by sequences that have high homology to a particular region of genomic DNA, two recombination events will yield the transfer of the foreign DNA into the genome (Figure 5.33). In this manner, specific genes can be targeted if their flanking nucleotide sequences are known.

For example, the gene-knockout approach has been applied to the genes encoding gene-regulatory proteins (also called *transcription factors*) that control the differentiation of muscle cells. When both copies of the gene for the regulatory protein *myogenin* are disrupted, an animal dies at birth because it lacks functional skeletal muscle. Microscopic inspection reveals that the tissues from which muscle normally forms contain precursor cells that have failed to differentiate fully (Figure 5.34A and B). Heterozygous mice containing one normal myogenin gene and one disrupted gene appear normal, suggesting that a reduced level of gene expression is still sufficient for normal muscle development. The generation and characterization of this knockout strain provided strong evidence that functional myogenin is essential for proper development of skeletal muscle tissue

FIGURE 5.33 Gene disruption by homologous recombination. (A) A mutated version of the gene to be disrupted is constructed, maintaining some regions of homology with the normal gene (red). When the foreign mutated gene is introduced into an embryonic stem cell, (B) recombination takes place at regions of homology and (C) the normal (targeted) gene is replaced, or "knocked out," by the foreign gene. The cell is inserted into embryos, and mice lacking the gene (knockout mice) are produced.



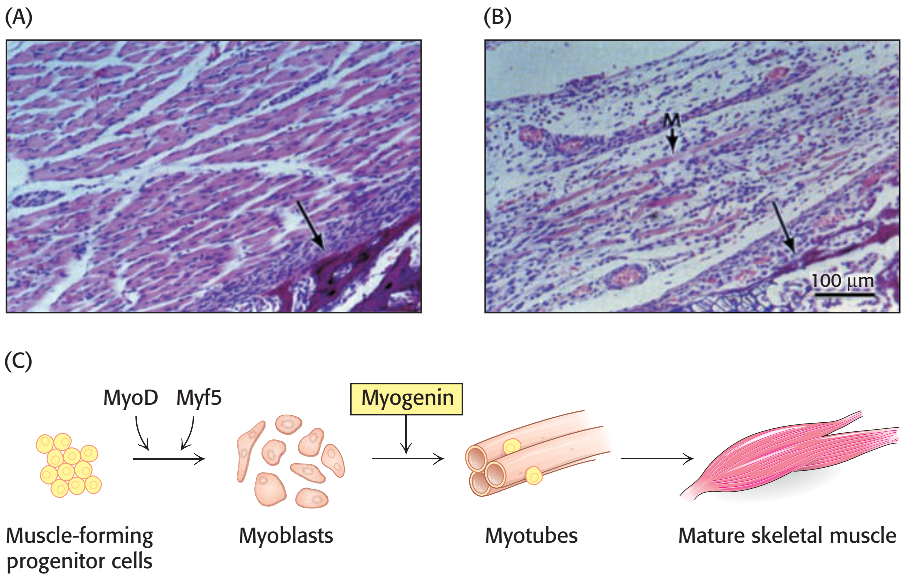


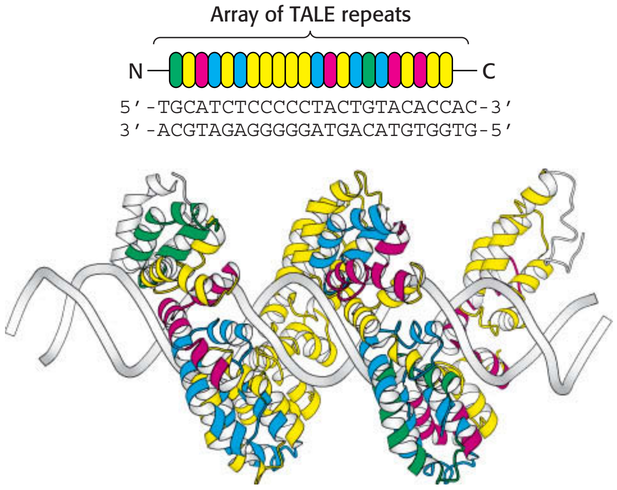
FIGURE 5.34 Consequences of gene disruption. Sections of muscle from normal (A) and myogenin-knockout (B) mice, as viewed under the light microscope. The unlabeled arrows in both panels identify comparable sections of the pelvic bone, indicating that similar anatomical regions are depicted. Muscles do not develop properly in mice having both myogenin genes disrupted. A poorly formed muscle fiber in the knockout strain is indicated by the M arrow. (C) The development of mature skeletal muscle from progenitor cells is a highly regulated process involving a number of intermediate cell types and multiple transcription factors. Through the gene-disruption studies in (A) and (B), myogenin was identified as an essential component of this pathway. [(A) and (B) From P. Hasty, et al., *Nature* 364:501–506, 1993; (C) Information from S. Hettmer and A. J. Wagers, *Nat. Med*. 16:171–173, 2010, Fig. 1]

(Figure 5.34C). Analogous studies have probed the function of many other genes to generate animal models for known human genetic diseases.

Manipulation of genomic DNA using homologous recombination, while a powerful tool, has limitations. Introduction of point mutations into genes, rather than knocking out the entire gene, can be difficult and time-consuming. In addition, these methods are generally limited to specific model organisms, such as yeast, mice, and fruit flies. Over the past 10 years, new methods for the highly specific modification of genomic DNA, or *genome editing*, have emerged. These approaches rely on engineered site-specific nucleases that introduce double-strand breaks at precisely-determined sequences within genomic DNA. In one approach, the nonspecific nuclease domain of the restriction enzyme FokI is fused to a DNA-binding domain designed to bind to a particular DNA sequence. In *zinc-finger nucleases* (ZFNs), the DNA-binding domain contains a series of zinc finger domains (Section 32.2), small zinc-binding motifs that each recognize a sequence of three base pairs. The preferred DNA binding sequence can be altered by changing the identity of only four contact residues within each finger. In *transcription activator-like effector nucleases* (TALENs), the DNA-binding domain is comprised of an array of TALE repeats. Each repeat contains 34 amino acids and two α-helices, yet only two of these residues (at positions 12 and 13) are responsible for the unique recognition of a single nucleotide within the double helix (Figure 5.35). Mutation of these residues within an array of repeats enables the recognition of a vast number of possible DNA target sequences with a high degree of specificity.

How do these engineered nucleases effect a change in the genomic DNA sequence? Upon binding of the ZFN or

FIGURE 5.35 TALE repeats recognize individual bases in DNA. Each TALE repeat contains 34 amino acids, two of which specify its nucleotide binding partner. In this figure, the identity of these residues is indicated by the color of the repeat. TALE proteins can be designed to uniquely recognize extended oligonucleotide sequences. In this example, a 22 base-pair sequence is bound by a single TALE protein, the bacterial effector PthXo1. [Drawn from 3UGM.pdb]


Array of TALE repeats
N
C
5′-TGCATCTCCCCCTACTGTACACCAC-3′
3′-ACGTAGAGGGGGATGACATGTGGTG-5′

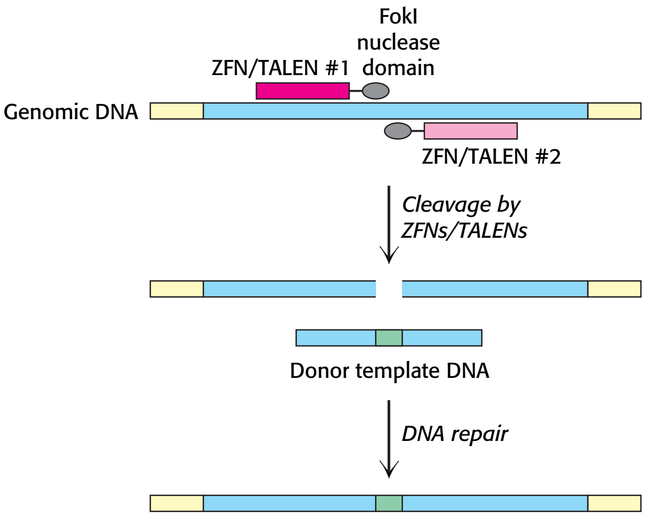


FIGURE 5.36 Genome editing by site-specific nucleases. A pair of ZFNs or TALENs cleave opposite strands of a targeted gene (blue) within the genome. The DNA repair machinery of the cell will use a homologous donor template DNA fragment to fix the double-strand break, incorporating the desired modifications (green) into the targeted gene.

TALEN to DNA, the nuclease domain cleaves the phosphate backbone of one of the DNA strands. A second nuclease, designed to recognize the opposite strand, introduces a second cleavage site, yielding a complete double-stranded break. The resulting cleavage site is repaired by the DNA repair machinery of the host cell (Section 28.5). If a DNA fragment containing the desired sequence change is simultaneously introduced with the nucleases, the repair machinery will use this donor template to introduce these changes directly into the genomic sequence, in a manner similar to the homologous recombination process described above (Figure 5.36).

Site-specific nuclease-based genome editing methods have now been applied to a variety of species, including model organisms used in the laboratory (rat, zebrafish, and fruit fly), various forms of livestock (pig, cow), and a number of plants. In addition, their use as therapeutic tools in humans is currently under investigation. For example, a ZFN which inactivates the human CCR5 gene, a coreceptor for cellular invasion of human immunodeficiency virus (HIV), is currently in clinical trials for the treatment of patients infected with HIV.

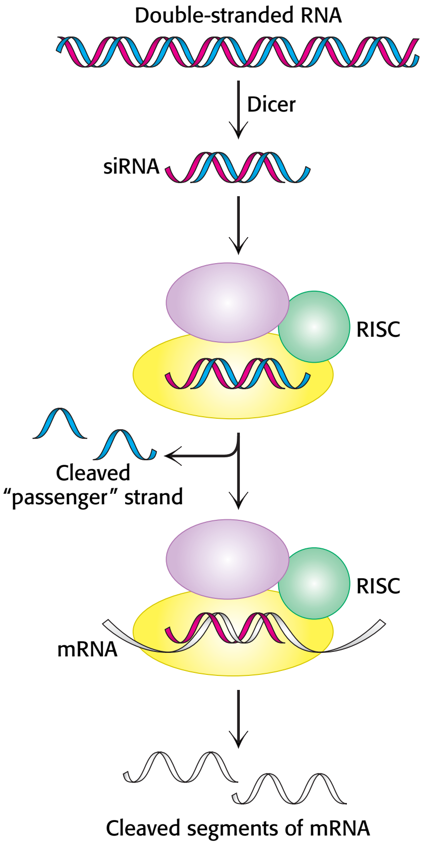


FIGURE 5.37 RNA interference mechanism. A double-stranded RNA molecule is cleaved into 21-bp fragments by the enzyme Dicer to produce siRNAs. These siRNAs are incorporated into the RNA-induced silencing complex (RISC), where the single-stranded RNAs guide the cleavage of mRNAs that contain complementary sequences.

RNA interference provides an additional tool for disrupting gene expression

An extremely powerful tool for disrupting gene expression was serendipitously discovered in the course of studies that required the introduction of RNA into a cell. The introduction of a specific double-stranded RNA molecule into a cell was found to suppress the transcription of genes that contained sequences present in the double-stranded RNA molecule. Thus, the introduction of a specific RNA molecule can interfere with the expression of a specific gene.

The mechanism of *RNA interference* has been largely established (Figure 5.37). When a double-stranded RNA molecule is introduced into an appropriate cell, the RNA is cleaved by the enzyme Dicer into fragments approximately 21 nucleotides in length. Each fragment, termed a small interfering RNA (siRNA), consists of 19 bp of double-stranded RNA and 2 bases of unpaired RNA on each 5′ end. The siRNA is loaded into an assembly of several proteins referred to as the *RNA-induced silencing complex* (RISC), which unwinds the RNA duplex and cleaves one of the strands, the so-called *passenger strand*. The uncleaved single-stranded RNA segment, the *guide strand*, remains incorporated into the enzyme. The fully assembled RISC cleaves mRNA molecules that contain exact complements of the guide-strand sequence. Thus, levels of such mRNA molecules are dramatically reduced. The technique of RNA interference is called *gene knockdown,* because the expression of the gene is reduced but not eliminated, as is the case with gene knockouts.

The machinery necessary for RNA interference is found in many cells. In some organisms such as *C. elegans,* RNA interference is quite efficient. Indeed, RNA interference can be induced simply by feeding *C. elegans* strains of *E. coli* that have been engineered to produce appropriate double-stranded RNA molecules. Although not as efficient in mammalian cells, RNA interference has emerged as a powerful research tool for reducing the expression of specific genes. Moreover, initial clinical trials of therapies based on RNA interference are underway.

Tumor-inducing plasmids can be used to introduce new genes into plant cells

The common soil bacterium *Agrobacterium tumefaciens* infects plants and introduces foreign genes into plant cells (Figure 5.38). A lump of tumor tissue called a *crown gall* grows at the site of infection. Crown galls synthesize opines, a group of amino acid derivatives that are metabolized by the infecting bacteria. In essence, the metabolism of the plant cell is diverted to satisfy the highly distinctive appetite of the intruder. *Tumor-inducing plasmids* (Ti plasmids) that are carried by *A. tumefaciens* carry instructions for the switch to the tumor state and the synthesis of opines. A small part of the Ti plasmid becomes integrated into the genome of infected plant cells; this 20-kb segment is called *T-DNA* (transferred DNA; Figure 5.39).

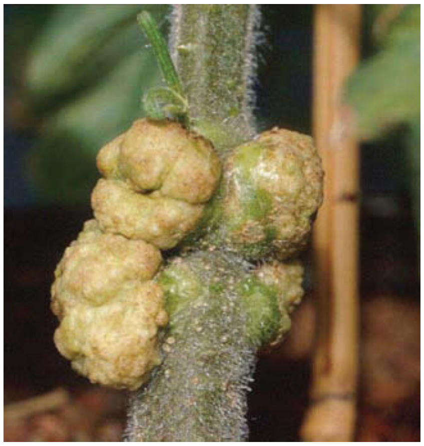


FIGURE 5.38 Tumors in plants. Crown gall, a plant tumor, is caused by a bacterium (*Agrobacterium tumefaciens*) that carries a tumor-inducing plasmid (Ti plasmid). [From M. Escobar et al., *Proc. Natl. Acad. Sci. U. S. A.* 98:13437–13442, 2001. Copyright © 2001 National Academy of Sciences, U. S. A.]

Ti-plasmid derivatives can be used as vectors to deliver foreign genes into plant cells. First, a segment of foreign DNA is inserted into the T-DNA region of a small plasmid through the use of restriction enzymes and ligases. This synthetic plasmid is added to *A. tumefaciens* colonies harboring naturally occurring Ti plasmids. By recombination, Ti plasmids containing the foreign gene are formed. These Ti vectors hold great promise as tools for exploring the genomes of plant cells and modifying plants to improve their agricultural value and crop yield. However, they are not suitable for transforming all types of plants. Ti-plasmid transfer is effective with dicots (broad-leaved plants such as grapes) and a few kinds of monocots but not as effective with economically important cereal monocots.

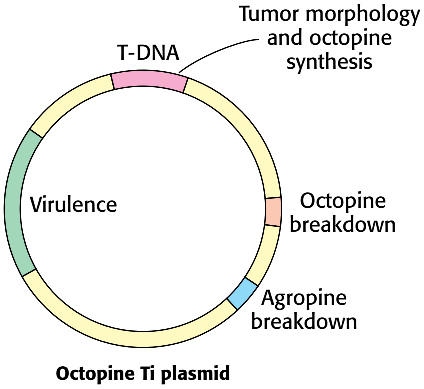
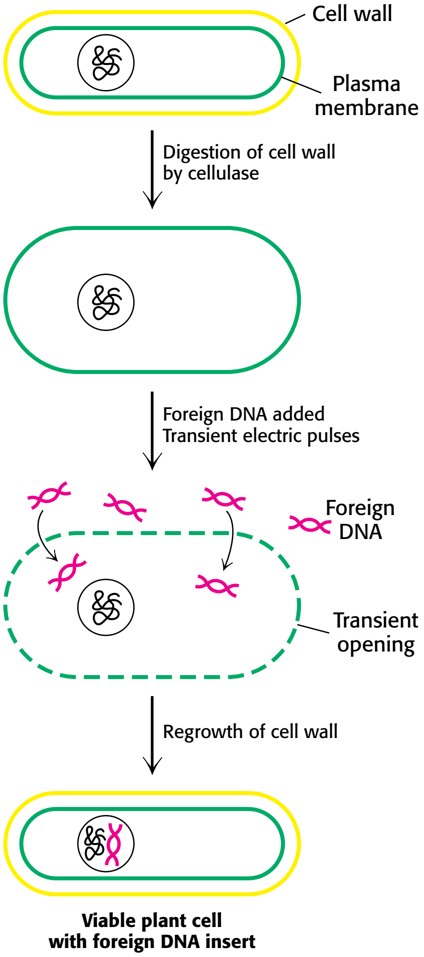


FIGURE 5.39 Ti plasmids. Agrobacteria containing Ti plasmids can deliver foreign genes into some plant cells. [Information from M. Chilton. A vector for introducing new genes into plants. Copyright © 1983 by Scientific American, Inc. All rights reserved.]

Foreign DNA can be introduced into cereal monocots as well as dicots by applying intense electric fields, a technique called *electroporation* (Figure 5.40). First, the cellulose wall surrounding plant cells is removed by adding cellulase; this treatment produces *protoplasts,* plant cells with exposed plasma membranes. Electric pulses are then applied to a suspension of protoplasts and plasmid DNA. Because high electric fields make membranes transiently permeable to large molecules, plasmid DNA molecules enter the cells. The cell wall is then allowed to reform, and the plant cells are again viable. Maize cells and carrot cells have been stably transformed in this way with the use of plasmid DNA that includes genes for resistance to herbicides. Moreover, the transformed cells efficiently express the plasmid DNA. Electroporation is also an effective means of delivering foreign DNA into animal cells and bacterial cells.

The most effective means of transforming plant cells is through the use of *"gene guns,"* or *bombardment-mediated transformation*. DNA is coated onto 1-μm-diameter tungsten pellets, and these microprojectiles are fired at the target cells with a velocity greater than 400 m s^{-1}. Despite its apparent crudeness, this technique is proving to be the most effective way of transforming



Cell wall
Plasma membrane
Digestion of cell wall by cellulase
Foreign DNA added
Transient electric pulses
Foreign DNA
Transient opening
Regrowth of cell wall
Viable plant cell with foreign DNA insert

FIGURE 5.40 Electroporation. Foreign DNA can be introduced into plant cells by electroporation, the application of intense electric fields to make their plasma membranes transiently permeable.

plants, especially important crop species such as soybean, corn, wheat, and rice. The gene-gun technique affords an opportunity to develop genetically modified organisms (GMOs) with beneficial characteristics, such as the ability to grow in poor soils, resistance to natural climatic variation, resistance to pests, and the ability to fortify nutritional content. These crops might be most useful in developing countries. However, the use of GMOs is highly controversial, as some fear that their safety risks have not been adequately addressed.

The first GMO to come to market was a tomato characterized by delayed ripening, rendering it ideal for shipment. Pectin is a polysaccharide that gives tomatoes their firmness and is naturally destroyed by the enzyme *polygalacturonase*. As pectin is destroyed, the tomatoes soften, making shipment difficult. DNA was introduced that disrupts the polygalacturonase gene. Less of the enzyme was produced, and the tomatoes stayed fresh longer. However, the tomato's poor taste hindered its commercial success. An especially successful result of the use of Ti plasmid to modify crops is golden rice. Golden rice is a variety of genetically modified rice that contains the genes for β-carotene synthesis, a required precursor for vitamin A synthesis in humans. Consumption of this rice will benefit children and pregnant woman in parts of the world where rice is a dietary staple and vitamin A deficiency is common.

Human gene therapy holds great promise for medicine

The field of *gene therapy* attempts to express specific genes within the human body in such a way that beneficial results are obtained. The gene targeted for expression may be already present or specially introduced. Alternatively, gene therapy may attempt to modify genes containing sequence variations that have harmful consequences. A tremendous amount of research remains to be done before gene therapy becomes practical. Nonetheless, considerable progress has been made. For example, some people lack functional genes for *adenosine deaminase* and succumb to infections if exposed to a normal environment, a condition called *severe combined immunodeficiency* (SCID). Functional genes for this enzyme have been introduced by using gene-therapy vectors based on retroviruses. Although these vectors have produced functional enzymes and reduced the clinical symptoms, challenges remain. These challenges include increasing the longevity of the effects and eliminating unwanted side effects. Future research promises to transform gene therapy into an important tool for clinical medicine.

SUMMARY

5.1 The Exploration of Genes Relies on Key Tools

The recombinant DNA revolution in biology is rooted in the repertoire of enzymes that act on nucleic acids. Restriction enzymes are a key group among them. These endonucleases recognize specific base sequences in double-helical DNA and cleave both strands of the duplex, forming specific fragments of DNA. These restriction fragments can be separated and displayed by gel electrophoresis. The pattern of these fragments on the gel is a fingerprint of a DNA molecule. A DNA fragment containing a particular sequence can be identified by hybridizing it with a labeled single-stranded DNA probe (Southern blotting).

Rapid sequencing techniques have been developed to further the analysis of DNA molecules. DNA can be sequenced by controlled interruption of replication. The fragments produced are separated by gel electrophoresis and visualized by autoradiography of a ^{32}P label at the 5′ end or by fluorescent tags.

DNA probes for hybridization reactions, as well as new genes, can be synthesized by the automated solid-phase method. DNA chains as long as 100 nucleotides can be readily synthesized. The polymerase chain reaction makes it possible to greatly amplify specific segments of DNA in vitro. The region amplified is determined by the placement of a pair of primers that are added to the target DNA along with a thermostable DNA polymerase and deoxyribonucleoside triphosphates. The exquisite sensitivity of PCR makes it a choice technique in detecting pathogens and cancer markers, in genotyping, and in amplifying DNA from fossils that are many thousands of years old.

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

New genes can be constructed in the laboratory, introduced into host cells, and expressed. Novel DNA molecules are made by joining fragments that have complementary cohesive ends produced by the action of a restriction enzyme. DNA ligase seals breaks in DNA chains. Vectors for propagating the DNA include plasmids, λ phage, and bacterial and yeast artificial chromosomes. Specific genes can be cloned from a genomic library with the use of a DNA or RNA probe. Foreign DNA can be expressed after insertion into prokaryotic and eukaryotic cells by the appropriate vector. Specific mutations can be generated in vitro to engineer novel proteins. A mutant protein with a single amino acid substitution can be produced by priming DNA replication with an oligonucleotide encoding the new amino acid. Plasmids can be engineered to permit the facile insertion of a DNA cassette containing any desired mutation. The techniques of protein and nucleic acid chemistry are highly synergistic. Investigators now move back and forth between gene and protein with great facility.

5.3 Complete Genomes Have Been Sequenced and Analyzed

The sequences of many important genomes are known in their entirety. More than 10,000 bacterial and archaeal genomes have been sequenced, including those from key model organisms and important pathogens. The sequence of the human genome has now been completed with nearly full coverage and high precision. Only from 20,000 to 25,000 protein-encoding genes appear to be present in the human genome, a substantially smaller number than earlier estimates. Comparative genomics has become a powerful tool for analyzing individual genomes and for exploring evolution. Genomewide gene-expression patterns can be examined through the use of DNA microarrays.

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

Changes in gene expression can be readily determined by such techniques as quantitative PCR and hybridization to microarrays. The production of transgenic mice carrying mutations known to cause ALS in humans has been a source of considerable insight into the disease mechanism and its possible treatment. The functions of particular genes can be investigated by disruption. One method of disrupting the expression of a particular gene is through RNA interference, which depends on the introduction of specific double-stranded RNA molecules into eukaryotic cells. New DNA can be brought into plant cells by the soil bacterium *Agrobacterium tumefaciens,* which harbors Ti plasmids. DNA can also be introduced into plant cells by applying intense electric fields, which render them transiently permeable to very large molecules, or by bombarding them with DNA-coated microparticles. Gene therapy holds great promise for clinical medicine, but many challenges remain.

KEY TERMS

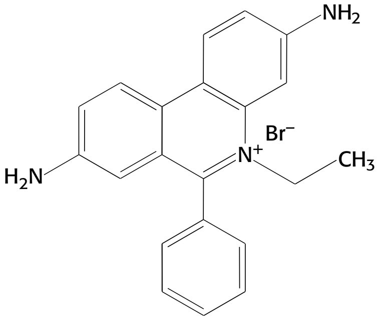
restriction enzyme (p. 137)
palindrome (p. 137)
DNA probe (p. 138)
Southern blotting (p. 138)
northern blotting (p. 138)
controlled termination of replication (Sanger dideoxy method) (p. 138)
polymerase chain reaction (PCR) (p. 141)
polymorphism (p. 143)
vector (p. 143)
plasmid (p. 144)
sticky ends (p. 144)
DNA ligase (p. 144)
cloning vector (p. 145)
reporter gene (p. 145)
expression vector (p. 146)
lambda (λ) phage (p. 146)
bacterial artificial chromosome (BAC) (p. 147)
yeast artificial chromosome (YAC) (p. 147)
genomic library (p. 148)
complementary DNA (cDNA) (p. 149)
reverse transcriptase (p. 149)
cDNA library (p. 149)
site-directed mutagenesis (p. 150)
cassette mutagenesis (p. 151)
pseudogene (p. 154)
mobile genetic element (p. 154)
short interspersed elements (SINES) (p. 154)
long interspersed elements (LINES) (p. 154)
next-generation sequencing (p. 155)
quantitative PCR (qPCR) (p. 157)
transcriptome (p. 158)
DNA microarray (gene chip) (p. 158)
transgenic mouse (p. 160)
gene disruption (gene knockout) (p. 160)
genome editing (p. 161)
zinc-finger nuclease (ZFN) (p. 161)
transcription activator-like effector nuclease (TALEN) (p. 161)
RNA interference (p. 162)
RNA-induced silencing complex (RISC) (p. 162)
tumor-inducing plasmid (Ti plasmid) (p. 163)
gene gun (bombardment-mediated transformation) (p. 163)

PROBLEMS

1. *It's not the heat . . .* Why is *Taq* polymerase especially useful for PCR?

2. *The right template.* Ovalbumin is the major protein of egg white. The chicken ovalbumin gene contains eight exons separated by seven introns. Should ovalbumin cDNA or ovalbumin genomic DNA be used to form the protein in *E. coli?* Why?

3. *Handle with care.* Ethidium bromide is a commonly used stain for DNA molecules after separation by gel electrophoresis. The chemical structure of ethidium bromide is shown here. Based on this structure, suggest how this stain binds to DNA.



Ethidium bromide

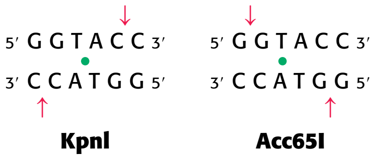
4. *Cleavage frequency.* The restriction enzyme AluI cleaves at the sequence 5′-AGCT-3′, and NotI cleaves at 5′-GCGGCCGC-3′. What would be the average distance between cleavage sites for each enzyme on digestion of double-stranded DNA? Assume that the DNA contains equal proportions of A, G, C, and T.

5. *The right cuts.* Suppose that a human genomic library is prepared by exhaustive digestion of human DNA with the EcoRI restriction enzyme. Fragments averaging about 4 kb in length would be generated. Is this procedure suitable for cloning large genes? Why or why not?

6. *A revealing cleavage.* Sickle-cell anemia arises from a mutation in the gene for the β chain of human hemoglobin. The change from GAG to GTG in the mutant eliminates a cleavage site for the restriction enzyme MstII, which recognizes the target sequence CCTGAGG. These findings form the basis of a diagnostic test for the sickle-cell gene. Propose a rapid procedure for distinguishing between the normal and the mutant gene. Would a positive result prove that the mutant contains GTG in place of GAG?

7. *Sticky ends?* The restriction enzymes KpnI and Acc65I recognize and cleave the same 6-bp sequence. However, the sticky end formed from KpnI cleavage cannot

be ligated directly to the sticky end formed from Acc65I cleavage. Explain why.



8. *Many melodies from one cassette.* Suppose that you have isolated an enzyme that digests paper pulp and have obtained its cDNA. The goal is to produce a mutant that is effective at high temperature. You have engineered a pair of unique restriction sites in the cDNA that flank a 30-bp coding region. Propose a rapid technique for generating many different mutations in this region.

9. *A blessing and a curse*. The power of PCR can also create problems. Suppose someone claims to have isolated dinosaur DNA by using PCR. What questions might you ask to determine if it is indeed dinosaur DNA?

10. *Rich or poor?* DNA sequences that are highly enriched in G–C base pairs typically have high melting temperatures. Moreover, once separated, single strands containing these regions can form rigid secondary structures. How might the presence of G–C-rich regions in a DNA template affect PCR amplification?

11. *Questions of accuracy*. The stringency of PCR amplification can be controlled by altering the temperature at which the primers and the target DNA undergo hybridization. How would altering the temperature of hybridization affect the amplification? Suppose that you have a particular yeast gene *A* and that you wish to see if it has a counterpart in humans. How would controlling the stringency of the hybridization help you?

12. *Terra incognita*. PCR is typically used to amplify DNA that lies between two known sequences. Suppose that you want to explore DNA on both sides of a single known sequence. Devise a variation of the usual PCR protocol that would enable you to amplify entirely new genomic terrain.

13. *A puzzling ladder.* A gel pattern displaying PCR products shows four strong bands. The four pieces of DNA have lengths that are approximately in the ratio of 1 : 2 : 3 : 4. The largest band is cut out of the gel, and PCR is repeated with the same primers. Again, a ladder of four bands is evident in the gel. What does this result reveal about the structure of the encoded protein?

14. *Chromosome walking*. Propose a method for isolating a DNA fragment that is adjacent in the genome to a previously isolated DNA fragment. Assume that you have access to a complete library of DNA fragments in a BAC vector but that the sequence of the genome under study has not yet been determined.

15. *Probe design.* Which of the following amino acid sequences would yield the most optimal oligonucleotide probe?

```
Ala-Met-Ser-Leu-Pro-Trp
Gly-Trp-Asp-Met-His-Lys
Cys-Val-Trp-Asn-Lys-Ile
Arg-Ser-Met-Leu-Gln-Asn
```

16. *Man's best friend.* Why might the genomic analysis of dogs be particularly useful for investigating the genes responsible for body size and other physical characteristics?

17. *Of mice and men.* You have identified a gene that is located on human chromosome 20 and wish to identify its location within the mouse genome. On which chromosome would you be most likely to find the mouse counterpart of this gene?

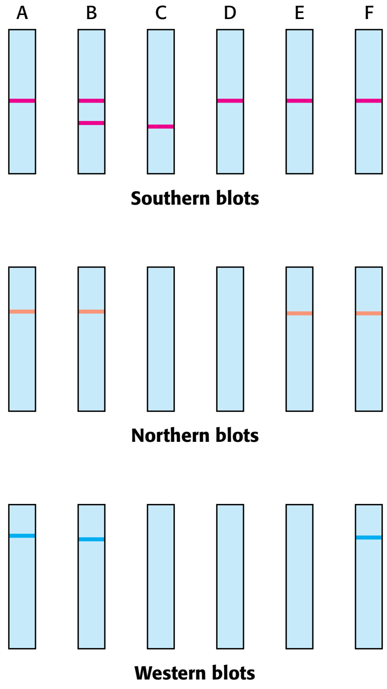
Chapter Integration Problems

18. *Designing primers I.* A successful PCR experiment often depends on designing the correct primers. In particular, the T_m for each primer should be approximately the same. What is the basis of this requirement?

19. *Designing primers II.* You wish to amplify a segment of DNA from a plasmid template by PCR with the use of the following primers: 5′-GGATCGATGCTCGCGA-3′ and 5′-AGGATCGGGTCGCGAG-3′. Despite repeated attempts, you fail to observe a PCR product of the expected length after electrophoresis on an agarose gel. Instead, you observe a bright smear on the gel with an approximate length of 25 to 30 base pairs. Explain these results.

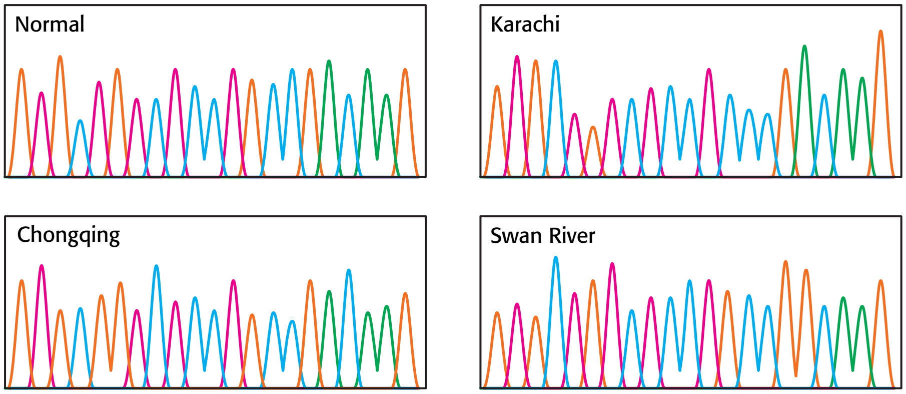
Chapter Integration and Data Interpretation Problem

20. *Any direction but east.* A series of people are found to have difficulty eliminating certain types of drugs from their bloodstreams. The problem has been linked to a gene *X*, which encodes an enzyme Y. Six people were tested with the use of various techniques of molecular biology. Person A is a normal control, person B is asymptomatic but some of his children have the metabolic problem, and persons C through F display the trait. Tissue samples from each person were obtained. Southern analysis was performed on the DNA after digestion with the restriction enzyme HindIII. Northern analysis of mRNA also was done. In both types of analysis, the gels were probed with labeled *X* cDNA. Finally, a western blot with an enzyme-linked monoclonal antibody was used to test for the presence of protein Y. The results are shown here. Why is person B without symptoms? Suggest possible defects in the other people.



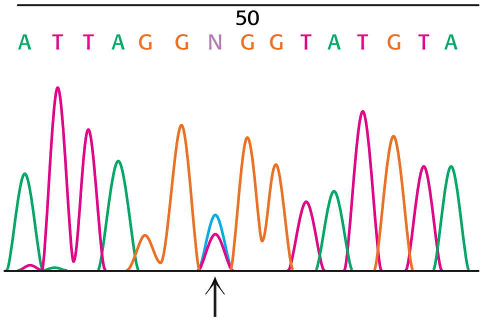
Data Interpretation Problems

21. *DNA diagnostics.* Representations of sequencing chromatograms for variants of the α chain of human hemoglobin are shown here. What is the nature of the amino acid change in each of the variants? The first triplet encodes valine.



Colors: ddATP, ddCTP, ddGTP, ddTTP

22. *Two peaks.* In the course of studying a gene and its possible mutation in humans, you obtain genomic DNA samples from a collection of persons and PCR amplify a region of interest within this gene. For one of the samples, you obtain the sequencing chromatogram shown here. Provide an explanation for the appearance of these data at position 49 (indicated by the arrow):



Animated Techniques

Visit www.whfreeman.com/Berg7e to see animations of Dideoxy Sequencing of DNA, Polymerase Chain Reaction, Synthesizing an Oligonucleotide Array, Screening an Oligonucleotide Array for Patterns of Gene Expression, Plasmid Cloning, In Vitro Mutagenesis of Cloned Genes, Creating a Transgenic Mouse. [Courtesy of H. Lodish et al., *Molecular Cell Biology*, 5th ed. (W. H. Freeman and Company, 2004).]

CHAPTER

6

Exploring Evolution and Bioinformatics

Evolutionary relationships are manifest in protein sequences. The close kinship between human beings and chimpanzees, hinted at by the mutual interest shown by Jane Goodall and a chimpanzee in the photograph, is revealed in the amino acid sequences of myoglobin. The human sequence (red) differs from the chimpanzee sequence (blue) in only one amino acid in a protein chain of 153 residues. [(Left) Kennan Ward/Corbis.]

GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL-
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL-

KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLHSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

Like members of a human family, members of molecular families often have features in common. Such family resemblance is most easily detected by comparing three-dimensional structure, the aspect of a molecule most closely linked to function. Consider as an example ribonuclease from cows, which was introduced in our consideration of protein folding (Section 2.6). Comparing structures reveals that the three-dimensional structure of this protein and that of a human ribonuclease are quite similar (Figure 6.1). Although the degree of overlap between these two structures is not unexpected, given their nearly identical biological functions, similarities revealed by other such comparisons are sometimes surprising. For example, angiogenin, a protein that stimulates the growth of new blood vessels, is also structurally similar to ribonuclease—so similar that both angiogenin and ribonuclease are clearly members of the same protein family (Figure 6.2). Angiogenin and ribonuclease must have had a common ancestor at some earlier stage of evolution.

Three-dimensional structures have been determined for only a small proportion of the total number of proteins. In contrast, gene sequences and the corresponding amino acid sequences are available for a great number of proteins, largely owing to the tremendous power of DNA cloning and sequencing techniques, including applications to complete-genome sequencing (Chapter 5). Evolutionary relationships also are manifest in amino acid sequences. For example, 35% of the amino acids in corresponding positions are identical in the sequences of bovine ribonuclease and angiogenin. Is this level sufficiently

OUTLINE

6.1 Homologs Are Descended from a Common Ancestor

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible

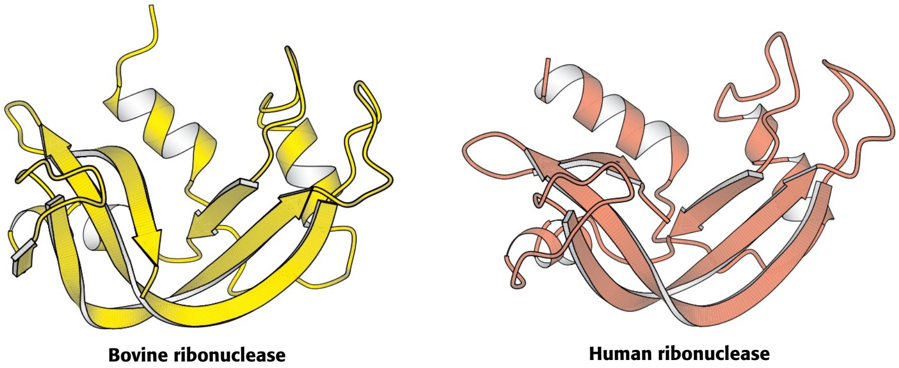


Figure 6.1 Structures of ribonucleases from cows and human beings. Structural similarity often follows functional similarity. [Drawn from 8RAT.pdb. and 2RNF.pdb.]

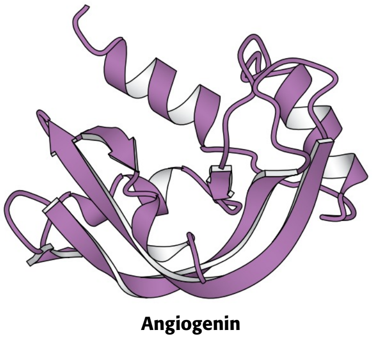


FIGURE 6.2 Structure of angiogenin. The protein angiogenin, identified on the basis of its ability to stimulate blood-vessel growth, is highly similar in three-dimensional structure to ribonuclease. [Drawn from 2ANG.pdb.]

high to ensure an evolutionary relationship? If not, what level is required? In this chapter, we shall examine the methods that are used to compare amino acid sequences and to deduce such evolutionary relationships.

Sequence-comparison methods are powerful tools in modern biochemistry. Sequence databases can be probed for matches to a newly elucidated sequence to identify related molecules. This information can often be a source of considerable insight into the function and mechanism of the newly sequenced molecule. When three-dimensional structures are available, they can be compared to confirm relationships suggested by sequence comparisons and to reveal others that are not readily detected at the level of sequence alone.

By examining the footprints present in modern protein sequences, the biochemist can learn about events in the evolutionary past. Sequence comparisons can often reveal pathways of evolutionary descent and estimated dates of specific evolutionary landmarks. This information can be used to construct evolutionary trees that trace the evolution of a particular protein or nucleic acid, in many cases from Archaea and Bacteria through Eukarya, including human beings. Molecular evolution can also be studied experimentally. In some cases, DNA from fossils can be amplified by PCR methods and sequenced, giving a direct view into the past. In addition, investigators can observe molecular evolution taking place in the laboratory, through experiments based on nucleic acid replication. The results of such studies are revealing more about how evolution proceeds.

6.1 Homologs Are Descended from a Common Ancestor

The exploration of biochemical evolution consists largely of an attempt to determine how proteins, other molecules, and biochemical pathways have been transformed through time. The most fundamental relationship between two entities is *homology;* two molecules are said to be *homologous* if they have been derived from a common ancestor. Homologous molecules, or *homologs,* can be divided into two classes (Figure 6.3). *Paralogs* are homologs that are present within one species. Paralogs often differ in their detailed biochemical functions. *Orthologs* are homologs that are present within different species and have very similar or identical functions. Understanding the homology between molecules can reveal the evolutionary history of the molecules as well as information about their function; if a newly sequenced protein is homologous to an already characterized protein, we have a strong indication of the new protein's biochemical function.

How can we tell whether two human proteins are paralogs or whether a yeast protein is the ortholog of a human protein? As will be discussed in Section 6.2, *homology is often detectable by significant similarity in nucleotide or amino acid sequence and almost always manifested in three-dimensional structure.*

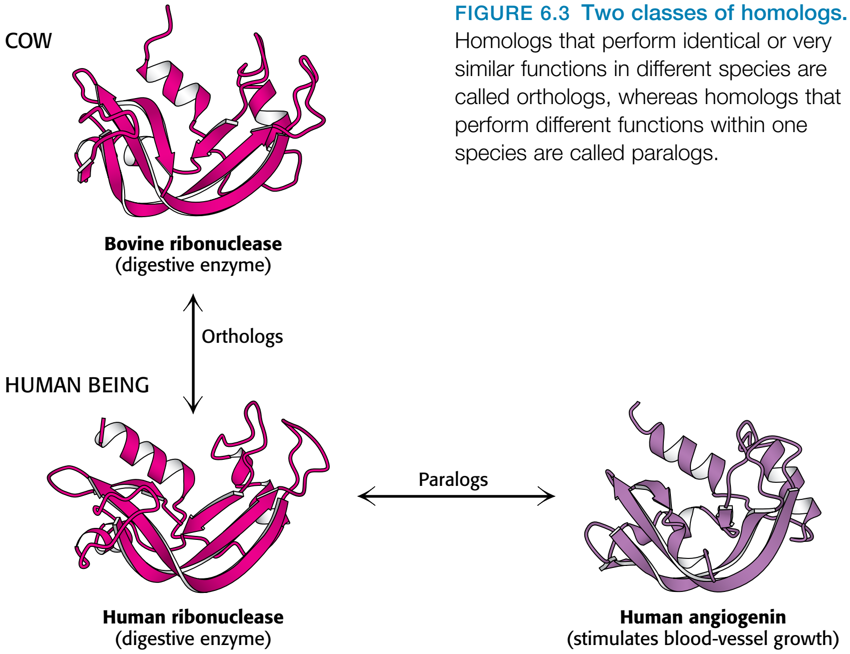


FIGURE 6.3 Two classes of homologs. Homologs that perform identical or very similar functions in different species are called orthologs, whereas homologs that perform different functions within one species are called paralogs.

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology

A significant sequence similarity between two molecules implies that they are likely to have the same evolutionary origin and, therefore, similar three-dimensional structures, functions, and mechanisms. Both nucleic acid and protein sequences can be compared to detect homology. However, the possibility exists that the observed agreement between any two sequences is solely a product of chance. Because nucleic acids are composed of fewer building blocks than proteins (4 bases versus 20 amino acids), the likelihood of random agreement between two DNA or RNA sequences is significantly greater than that for protein sequences. For this reason, detection of homology between protein sequences is typically far more effective.

To illustrate sequence-comparison methods, let us consider a class of proteins called the *globins.* Myoglobin is a protein that binds oxygen in muscle, whereas hemoglobin is the oxygen-carrying protein in blood (Chapter 7). Both proteins cradle a heme group, an iron-containing organic molecule that binds the oxygen. Each human hemoglobin molecule is composed of four heme-containing polypeptide chains, two identical α chains and two identical β chains. Here, we shall consider only the α chain. To examine the similarity between the amino acid sequence of the human α chain and that of human myoglobin (Figure 6.4), we apply a method, referred to as a *sequence alignment,* in which the two sequences are systematically aligned with respect to each other to identify regions of significant overlap.

Human hemoglobin (α chain)

```
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHG
SAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLS
HCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
```

Human myoglobin

```
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKS
EDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVK
YLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```

FIGURE 6.4 Amino acid sequences of human hemoglobin (α chain) and human myoglobin. α-Hemoglobin is composed of 141 amino acids; myoglobin consists of 153 amino acids. (One-letter abbreviations designating amino acids are used; see Table 2.2.)

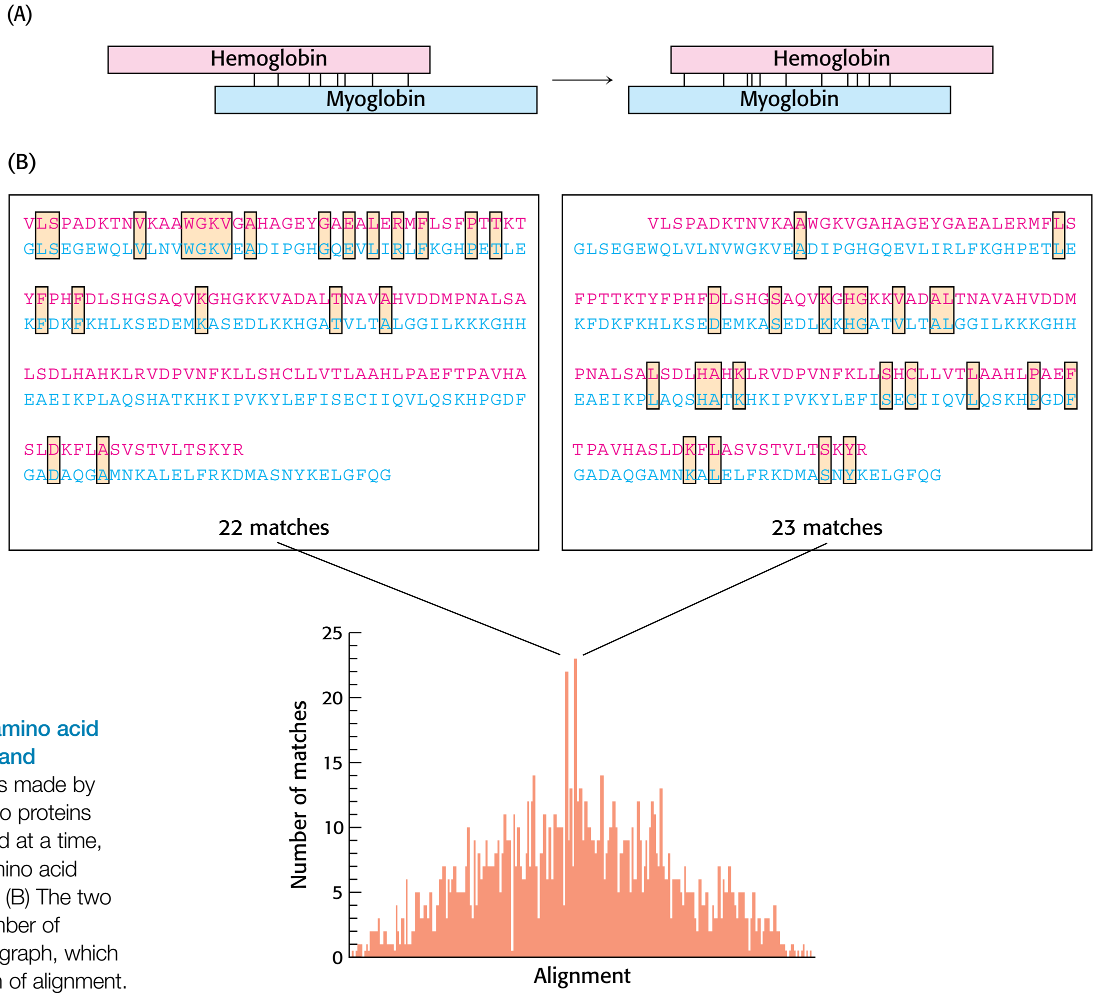
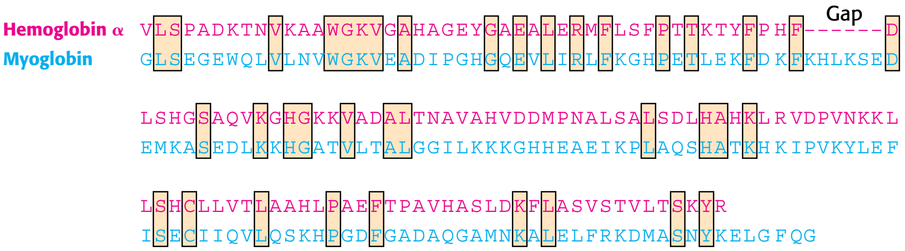


FIGURE 6.5 Comparing the amino acid sequences of α-hemoglobin and myoglobin. (A) A comparison is made by sliding the sequences of the two proteins past each other, one amino acid at a time, and counting the number of amino acid identities between the proteins. (B) The two alignments with the largest number of matches are shown above the graph, which plots the matches as a function of alignment.

How can we tell where to align the two sequences? In the course of evolution, the sequences of two proteins that have an ancestor in common will have diverged in a variety of ways. Insertions and deletions may have occurred at the ends of the proteins or within the functional domains themselves. Individual amino acids may have been mutated to other residues of varying degrees of similarity. To understand how the methods of sequence alignment take these potential sequence variations into account, let us first consider the simplest approach, where we slide one sequence past the other, one amino acid at a time, and count the number of matched residues, or *sequence identities* (Figure 6.5). For α-hemoglobin and myoglobin, the best alignment reveals 23 sequence identities, spread throughout the central parts of the sequences.

However, careful examination of all the possible alignments and their scores suggests that important information regarding the relationship between myoglobin and hemoglobin α has been lost with this method. In particular, we see that another alignment, featuring 22 identities, is nearly as good. This alignment is shifted by six residues relative to the preceding alignment and yields identities that are concentrated toward the amino-terminal end of the sequences. By introducing a *gap* into one of the sequences, the identities found in *both* alignments will be represented (Figure 6.6). Insertion of gaps allows the alignment method to compensate for the insertions or deletions of nucleotides that may have taken place in the gene for one molecule but not the other in the course of evolution.


Gap
Hemoglobin α
Myoglobin
VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF------D
GLSEGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSED
LSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNKKL
EMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEF
LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
ISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

38 identities: $38 \times (+10) = 380$
1 gap: $1 \times (-25) = \underline{-25}$
355

FIGURE 6.6 Alignment with gap insertion. The alignment of α-hemoglobin and myoglobin after a gap has been inserted into the hemoglobin α sequence.

The use of gaps substantially increases the complexity of sequence alignment because a vast number of possible gaps, varying in both position and length, must be considered throughout each sequence. Moreover, the introduction of an excessive number of gaps can yield an artificially high number of identities. Nevertheless, methods have been developed for the insertion of gaps in the automatic alignment of sequences. These methods use scoring systems to compare different alignments, including penalties for gaps to prevent the insertion of an unreasonable number of them. For example, in one scoring system, each identity between aligned sequences is counted as +10 points, whereas each gap introduced, regardless of size, counts for −25 points. For the alignment shown in Figure 6.6, there are 38 identities ($38 \times 10 = 380$) and 1 gap ($1 \times -25 = -25$), producing a score of ($380 + -25 = 355$). Overall, there are 38 matched amino acids in an average length of 147 residues; thus, the sequences are 25.9% identical. Next, we must determine the significance of this score and level of identity.

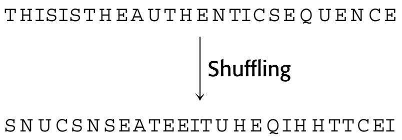

THISISTHEAUTHENTICSEQUENCE
Shuffling
SNUCSNSEATEEITUHEQIHHTTCEI

FIGURE 6.7 The generation of a shuffled sequence.

The statistical significance of alignments can be estimated by shuffling

The similarities in sequence in Figure 6.5 appear striking, yet there remains the possibility that a grouping of sequence identities has occurred by chance alone. Because proteins are composed of the same set of 20 amino acid monomers, the alignment of any two unrelated proteins will yield some identities, particularly if we allow the introduction of gaps. Even if two proteins have identical amino acid composition, they may not be linked by evolution. It is the order of the residues within their sequences that implies a relationship between them. Hence, we can assess the significance of our alignment by "shuffling," or randomly rearranging, one of the sequences (Figure 6.7), repeating the sequence alignment, and determining a new alignment score. This process is repeated many times to yield a histogram showing, for each possible score, the number of shuffled sequences that received that score (Figure 6.8). If the original score is not appreciably different from the scores from the shuffled alignments, then we cannot exclude the possibility that the original alignment is merely a consequence of chance.

When this procedure is applied to the sequences of myoglobin and α-hemoglobin, the authentic alignment (indicated by the red bar in Figure 6.8) clearly stands out. Its score is far above the mean for the alignment scores

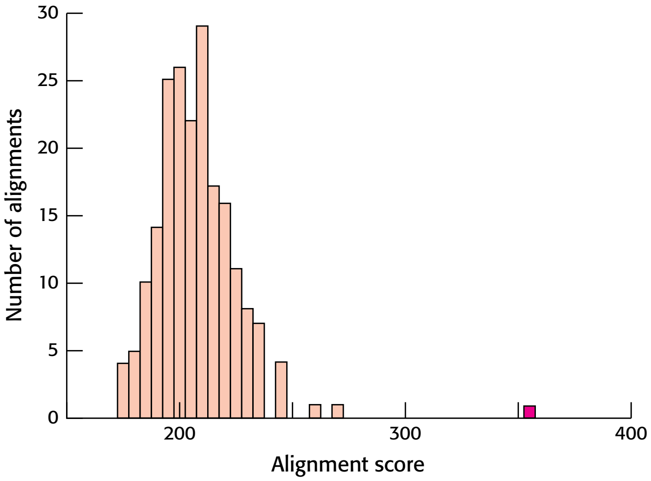


FIGURE 6.8 Statistical comparison of alignment scores. Alignment scores are calculated for many shuffled sequences, and the number of sequences generating a particular score is plotted against the score. The resulting plot is a distribution of alignment scores occurring by chance. The alignment score for unshuffled α-hemoglobin and myoglobin (shown in red) is substantially greater than any of these scores, strongly suggesting that the sequence similarity is significant.

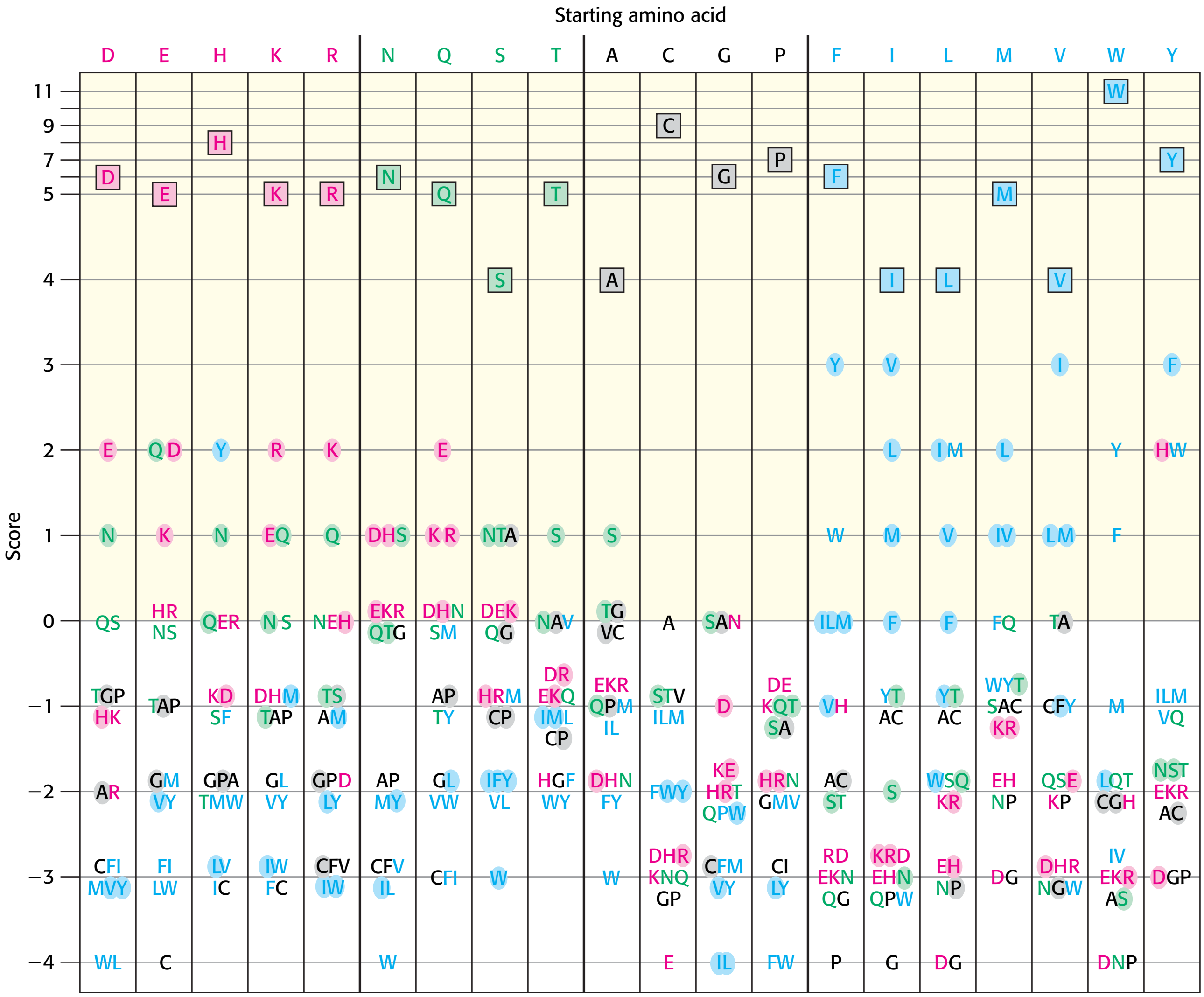
based on shuffled sequences. The probability that such a deviation occurred by chance alone is approximately 1 in 10^{20}. Thus, we can comfortably conclude that the two sequences are genuinely similar; the simplest explanation for this similarity is that these sequences are homologous—that is, the two molecules have descended from a common ancestor.

Distant evolutionary relationships can be detected through the use of substitution matrices

The scoring scheme heretofore described assigns points only to positions occupied by identical amino acids in the two sequences being compared. No credit is given for any pairing that is not an identity. However, as already discussed, two proteins related by evolution undergo amino acid substitutions as they diverge. A scoring system based solely on amino acid identity cannot account for these changes. To add greater sensitivity to the detection of evolutionary relationships, methods have been developed to compare two amino acids and assess their degree of *similarity*.

Not all substitutions are equivalent. For example, amino acid changes can be classified as structurally conservative or nonconservative. A *conservative substitution* replaces one amino acid with another that is similar in size and chemical properties. Conservative substitutions may have only minor effects on protein structure and often can be tolerated without compromising protein function. In contrast, in a *nonconservative substitution*, an amino acid is replaced by one that is structurally dissimilar. Amino acid changes can also be classified by the fewest number of nucleotide changes necessary to achieve the corresponding amino acid change. Some substitutions arise from the replacement of only a single nucleotide in the gene sequence; whereas others require two or three replacements. Conservative and single-nucleotide substitutions are likely to be more common than are substitutions with more radical effects.

How can we account for the type of substitution when comparing sequences? We can approach this problem by first examining the substitutions that have been observed in proteins known to be evolutionarily related. From an examination of appropriately aligned sequences, substitution matrices have been deduced. A *substitution matrix* describes a scoring system for the replacement of any amino acid with each of the other 19 amino acids. In these matrices, a large positive score corresponds to a substitution that occurs relatively frequently, whereas a large negative score corresponds to a substitution that occurs only rarely. A commonly used substitution matrix, the Blosum-62 (for *Blo*cks of amino acid *su*bstitution *m*atrix), is illustrated in Figure 6.9. In this depiction, each column in this matrix represents one of the 20 amino acids, whereas the position of the single-letter codes within each column specifies the score for the corresponding substitution. Notice that scores corresponding to identity (the boxed codes at the top of each column) are not the same for each residue, owing to the fact that less frequently occurring amino acids such as cysteine (C) and tryptophan (W) will align by chance less often than the more common residues. Furthermore, structurally conservative substitutions such as lysine (K) for arginine (R) and isoleucine (I) for valine (V) have relatively high scores, whereas nonconservative substitutions such as lysine for tryptophan result in negative scores (Figure 6.10). When two sequences are compared, each pair of aligned residues is assigned a score based on the matrix. In addition, gap penalties are often assessed. For example, the introduction of a single-residue gap lowers the alignment score by 12 points and the extension of an existing gap costs 2 points per residue. With the use of this scoring system, the alignment shown in Figure 6.6 receives a score of 115. In many regions, most substitutions are



Score	D	E	H	K	R	N	Q	S	T	A	C	G	P	F	I	L	M	V	W	Y
11																			W	
10																				
9											C									
8			H																	
7													P							Y
6	D					N						G		F						
5		E		K	R		Q		T								M			
4								S		A					I	L		V		
3														Y	V			I		F
2	E	QD	Y	R	K		E								L	IM	L		Y	HW
1	N	K	N	EQ	Q	DHS	KR	NTA	S	S				W	M	V	IV	LM	F	
0	QS	HR NS	QER	NS	NEH	EKR QTG	DHN SM	DEK QG	NAV	TG VC	A	SAN		ILM	F	F	FQ	TA		
−1	TGP HK	TAP	KD SF	DHM TAP	TS AM		AP TY	HRM CP	DR EKQ IML CP	EKR QPM IL	STV ILM	D	DE KQT SA	VH	YT AC	YT AC	WYT SAC KR	CFY	M	ILM VQ
−2	AR	GM VY	GPA TMW	GL VY	GPD LY	AP MY	GL VW	IFY VL	HGF WY	DHN FY	FWY	KE HRT QPW	HRN GMV	AC ST	S	WSQ KR	EH NP	QSE KP	LQT CGH	NST EKR AC
−3	CFI MVY	FI LW	LV IC	IW FC	CFV IW	CFV IL	CFI	W		W	DHR KNQ GP	CFM VY	CI LY	RD EKN QG	KRD EHN QPW	EH NP	DG	DHR NGW	IV EKR AS	DGP
−4	WL	C				W					E	IL	FW	P	G	DG			DNP	

FIGURE 6.9 A graphic view of the Blosum-62. This substitution matrix was derived by examining substitutions within aligned sequence blocks in related proteins. Amino acids are classified into four groups (charged, red; polar, green; large and hydrophobic, blue; other, black). Substitutions that require the change of only a single nucleotide are shaded. Identities are boxed. To find the score for a substitution of, for instance, a Y for an H, you find the Y in the column having H at the top and check the number at the left. In this case, the resulting score is 2.

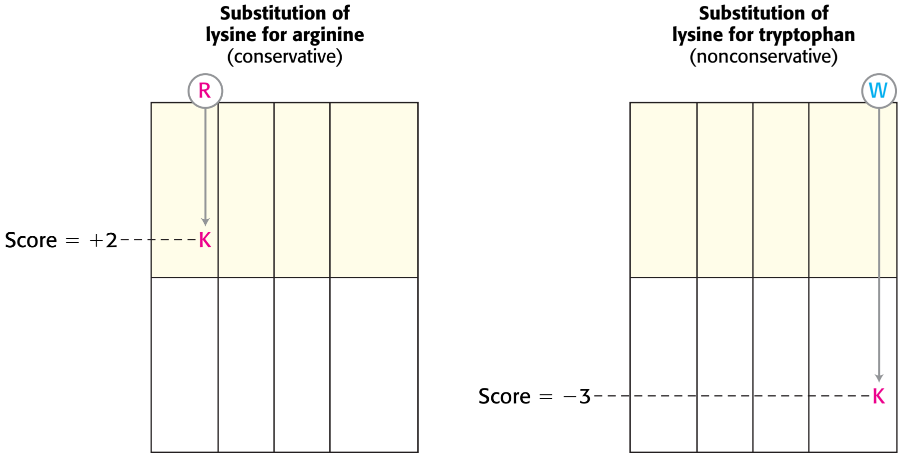
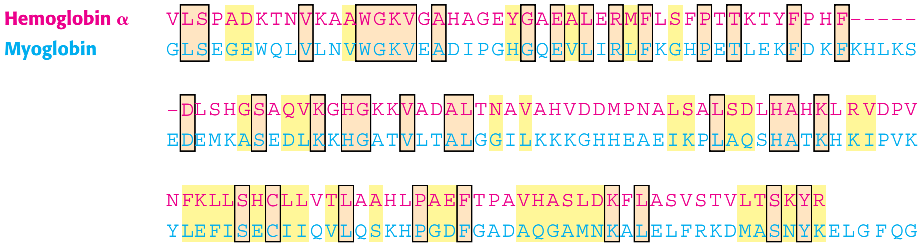


FIGURE 6.10 Scoring of conservative and nonconservative substitutions. The Blosum-62 indicates that a conservative substitution (lysine for arginine) receives a positive score, whereas a nonconservative substitution (lysine for tryptophan) is scored negatively. The matrix is depicted as an abbreviated form of Figure 6.9.



VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-----
GLSEGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKS

-DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPV
EDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVK

NFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
YLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

FIGURE 6.11 Alignment with conservative substitutions noted. The alignment of α-hemoglobin and myoglobin with conservative substitutions indicated by yellow shading and identities by orange.

conservative (defined as those substitutions with scores greater than 0) and relatively few are strongly disfavored (Figure 6.11).

This scoring system detects homology between less obviously related sequences with greater sensitivity than would a comparison of identities only. Consider, for example, the protein leghemoglobin, an oxygen-binding protein found in the roots of some plants. The amino acid sequence of leghemoglobin from the herb lupine can be aligned with that of human myoglobin and scored by using either the simple scoring scheme based on identities only or the Blosum-62 (Figure 6.9). Repeated shuffling and scoring provides a distribution of alignment scores (Figure 6.12). Scoring based solely on identities indicates that the probability of the alignment between myoglobin and leghemoglobin occurring by chance alone is 1 in 20. Thus, although the level of similarity suggests a relationship, there is a 5% chance that the similarity is accidental on the basis of this analysis. In contrast, users of the substitution matrix are able to incorporate the effects of conservative substitutions. From such an analysis, the odds of the alignment occurring by chance are calculated to be approximately 1 in 300. Thus, an analysis performed with the substitution matrix reaches a much firmer conclusion about the evolutionary relationship between these proteins (Figure 6.13).

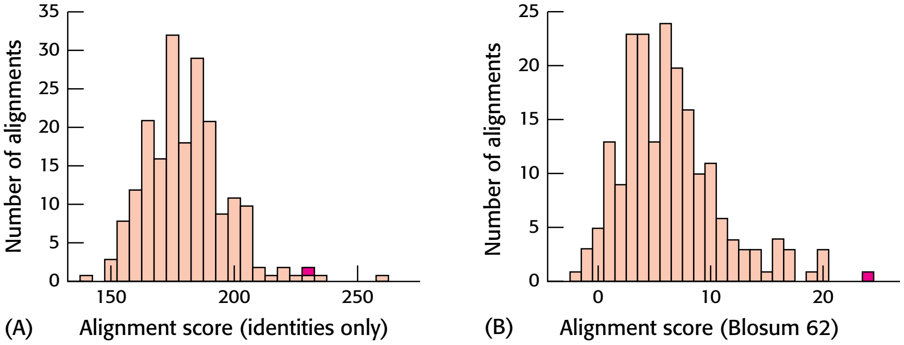
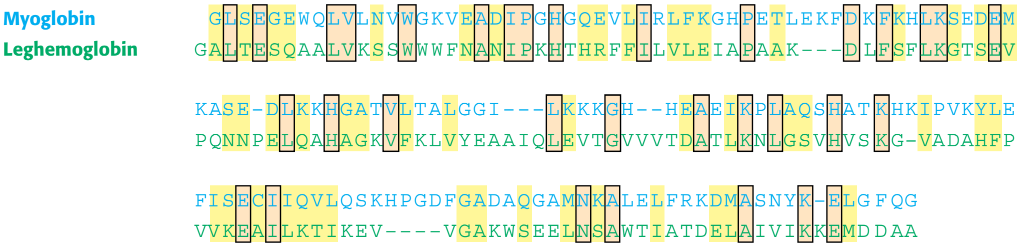


FIGURE 6.12 Alignment of identities only versus the Blosum-62. Repeated shuffling and scoring reveal the significance of sequence alignment for human myoglobin versus lupine leghemoglobin with the use of either (A) the simple, identity-based scoring system or (B) the Blosum-62. The scores for the alignment of the authentic sequences are shown in red. Accounting for amino acid similarity in addition to identity reveals a greater separation between the authentic alignment and the population of shuffled alignments.



GLSEGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEM
GALTESQAALVKSSWWWFNANIPKHTHRFFILVLEIAPAAK---DLFSFLKGTSEV

KASE-DLKKHGATVLTALGGI---LKKKGH--HEAEIKPLAQSHATKHKIPVKYLE
PQNNPELQAHAGKVFKLVYEAAIQLEVTGVVVTDATLKNLGSVHVSKG-VADAHFP

FISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYK-ELGFQG
VVKEAILKTIKEV----VGAKWSEELNSAWTIATDELAIVIKKEMDDAA

FIGURE 6.13 Alignment of human myoglobin and lupine leghemoglobin. The use of Blosum-62 yields the alignment shown between human myoglobin and lupine leghemoglobin, illustrating identities (orange boxes) and conservative substitutions (yellow). These sequences are 23% identical.

Experience with sequence analysis has led to the development of simpler rules of thumb. For sequences longer than 100 amino acids, sequence identities greater than 25% are almost certainly not the result of chance alone; such sequences are probably homologous. In contrast, if two sequences are less than 15% identical, their alignment alone is unlikely to indicate statistically significant similarity. For sequences that are between 15 and 25% identical, further analysis is necessary to determine the statistical significance of the alignment. It must be emphasized that *the lack of a statistically significant degree of sequence similarity does not rule out homology*. The sequences of many proteins that have descended from common ancestors have diverged to such an extent that the relationship between the proteins can no longer be detected from their sequences alone. As we will see, such homologous proteins can often be detected by examining three-dimensional structures.

Databases can be searched to identify homologous sequences

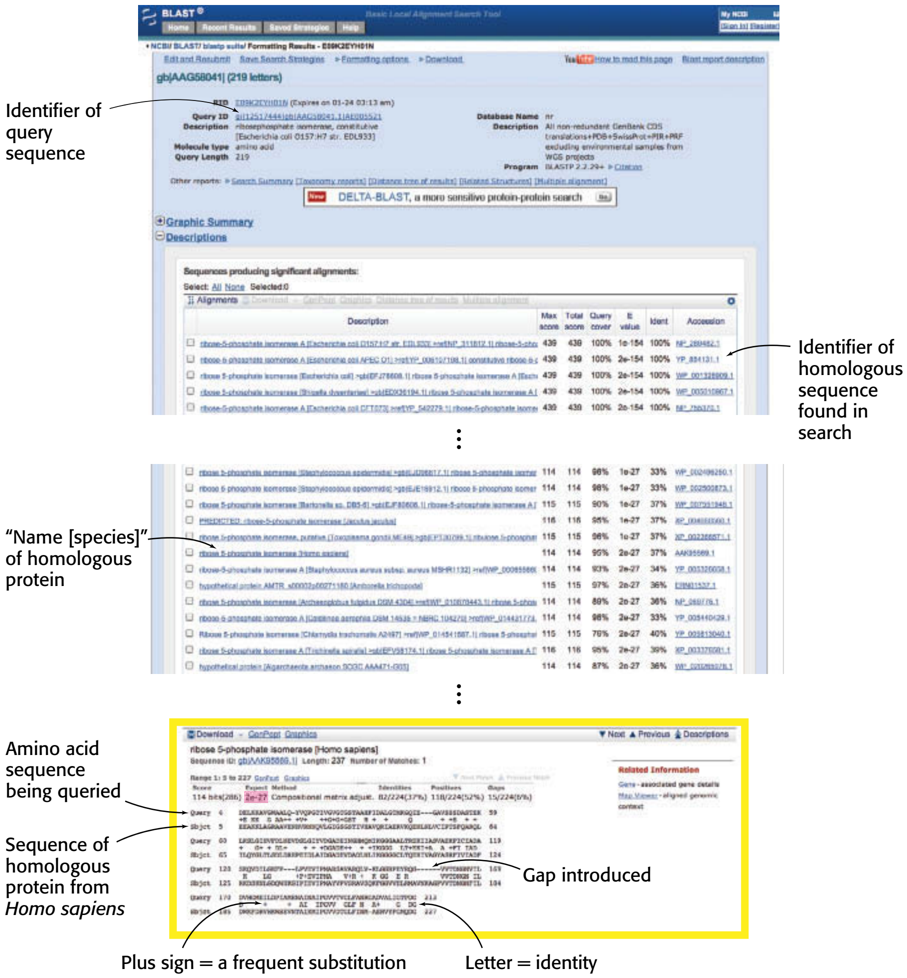
When the sequence of a protein is first determined, comparing it with all previously characterized sequences can be a source of tremendous insight into its evolutionary relatives and, hence, its structure and function. Indeed, an extensive sequence comparison is almost always the first analysis performed on a newly elucidated sequence. The sequence-alignment methods just described are used to compare an individual sequence with all members of a database of known sequences.

Database searches for homologous sequences are most often accomplished by using resources available on the Internet at the National Center for Biotechnology Information (www.ncbi.nih.gov). The procedure used is referred to as a *BLAST* (Basic Local Alignment Search Tool) *search*. An amino acid sequence is typed or pasted into the Web browser, and a search is performed, most often against a nonredundant database of all known sequences. At the end of 2013, this database included more than 35 million sequences. A BLAST search yields a list of sequence alignments, each accompanied by an estimate giving the likelihood that the alignment occurred by chance (Figure 6.14).

In 1995, investigators reported the first complete sequence of the genome of a free-living organism, the bacterium *Haemophilus influenzae*. With the sequences available, they performed a BLAST search with each deduced protein sequence. Of 1743 identified protein-coding regions, also called *open reading frames (ORFs)*, 1007 (58%) could be linked to some protein of known function that had been previously characterized in another organism. An additional 347 ORFs could be linked to sequences in the database for which no function had yet been assigned ("hypothetical proteins"). The remaining 389 sequences did not match any sequence present in the database at that time. Thus, investigators were able to identify likely functions for more than half the proteins within this organism solely by sequence comparisons.

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships

Sequence comparison is a powerful tool for extending our knowledge of protein function and kinship. However, biomolecules generally function as intricate three-dimensional structures rather than as linear polymers. Mutations occur at the level of sequence, but the effects of the mutations are


BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
NCBI/ BLAST/ blastp suite/ Formatting Results - E09K2EYH01N
Edit and Resubmit
Save Search Strategies
Formatting options
Download
How to read this page
Blast report description
DELTA-BLAST, a more sensitive protein-protein search
Graphic Summary
Descriptions
Download
Next
Previous
Descriptions
Related Information
Identifier of query sequence
Identifier of homologous sequence found in search
"Name [species]" of homologous protein
Amino acid sequence being queried
Sequence of homologous protein from *Homo sapiens*
Gap introduced
Plus sign = a frequent substitution
Letter = identity

gb|AAG58041| (219 letters)

RID E09K2EYH01N (Expires on 01-24 03:13 am)
Query ID gi|12517444|gb|AAG58041.1|AE005521
Description ribosephosphate isomerase, constitutive [Escherichia coli O157:H7 str. EDL933]
Molecule type amino acid
Query Length 219
Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.2.29+

Sequences producing significant alignments:

ribose 5-phosphate isomerase [Homo sapiens]
Sequence ID: gb|AAK95869.1| Length: 237 Number of Matches: 1

Range 1: 5 to 227

Score	Expect	Method	Identities	Positives	Gaps
114 bits(286)	2e-27	Compositional matrix adjust.	82/224(37%)	118/224(52%)	15/224(6%)

FIGURE 6.14 BLAST search results. Part of the results from a BLAST search of the nonredundant (nr) protein sequence database using the sequence of ribose 5-phosphate isomerase (also called phosphopentose isomerase, Chapter 20) from *E. coli* as a query. Among the thousands of sequences found is the orthologous sequence from humans, and the alignment between these sequences is shown (highlighted in yellow). The number of sequences with this level of similarity expected to be in the database by chance is 2×10^{-27} as shown by the E value (highlighted in red). Because this value is much less than 1, the observed sequence alignment is highly significant.

at the level of function, and function is directly related to tertiary structure. Consequently, to gain a deeper understanding of evolutionary relationships between proteins, we must examine three-dimensional structures, especially in conjunction with sequence information. The techniques of structural determination were presented in Section 3.5.

Tertiary structure is more conserved than primary structure

Because three-dimensional structure is much more closely associated with function than is sequence, tertiary structure is more evolutionarily conserved than is primary structure. This conservation is apparent in the tertiary structures of the globins (Figure 6.15), which are extremely similar even though the similarity between human myoglobin and lupine leghemoglobin is just barely detectable at the sequence level (Figure 6.12) and that between human α-hemoglobin and lupine leghemoglobin is not statistically significant (15% identity). This structural similarity firmly establishes that the framework that binds the heme group and facilitates the reversible binding of oxygen has been conserved over a long evolutionary period.

Anyone aware of the similar biochemical functions of hemoglobin, myoglobin, and leghemoglobin could expect the structural similarities. In a growing number of other cases, however, a comparison of three-dimensional structures has revealed striking similarities between proteins that were *not* expected to be related, on the basis of their diverse functions. A case in point is the protein actin, a major component of the cytoskeleton (Section 35.2), and heat shock protein 70 (Hsp70), which assists protein folding inside cells. These two proteins were found to be noticeably similar in structure despite only 16% sequence identity (Figure 6.16). On the basis of their three-dimensional structures, actin and Hsp70 are paralogs. The level of structural similarity strongly suggests that, despite their different biological roles in modern organisms, these proteins descended from a common ancestor. As the three-dimensional structures of more proteins are determined, such unexpected kinships are being discovered with increasing frequency. The search for such kinships relies ever more frequently on computer-based searches that are able to compare the three-dimensional structure of any protein with all other known structures.

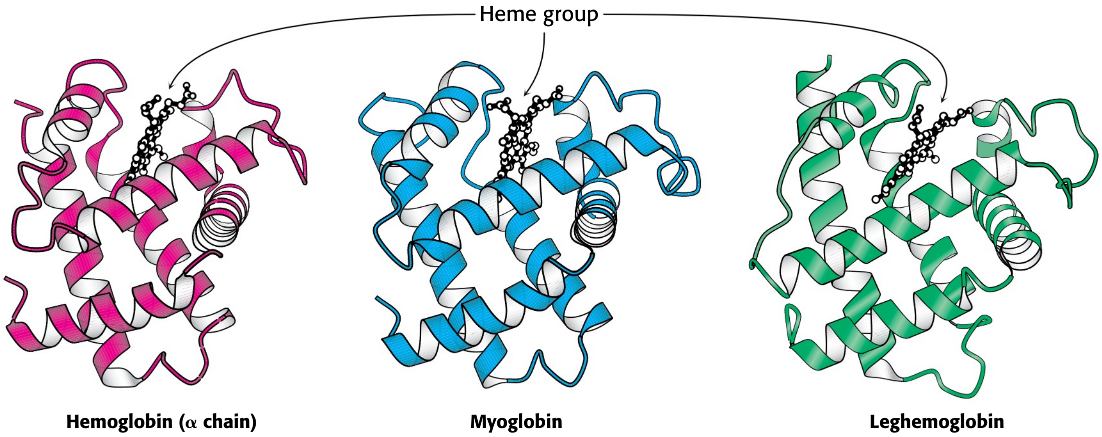


FIGURE 6.15 Conservation of three-dimensional structure. The tertiary structures of human hemoglobin (α chain), human myoglobin, and lupine leghemoglobin are conserved. Each heme group contains an iron atom to which oxygen binds. [Drawn from 1HBB.pdb, 1MBD.pdb, and 1GDJ.pdb.]

Knowledge of three-dimensional structures can aid in the evaluation of sequence alignments

The sequence-comparison methods described thus far treat all positions within a sequence equally. However, we know from examining families

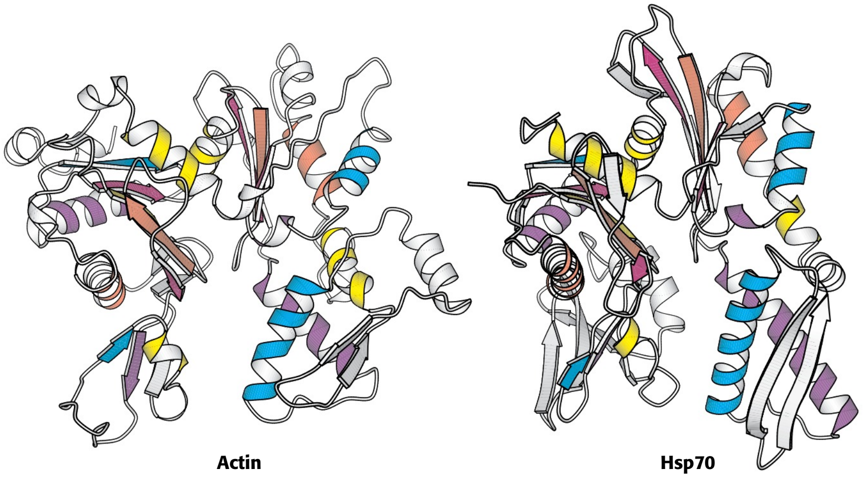


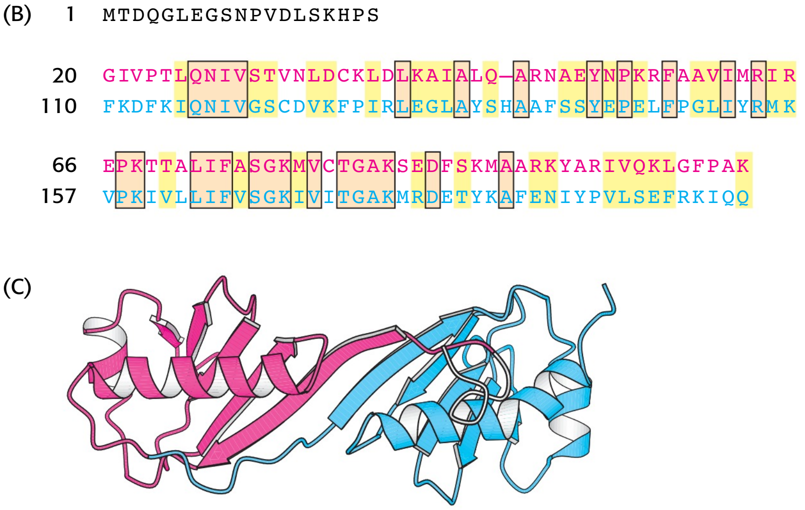
FIGURE 6.16 Structures of actin and a large fragment of heat shock protein 70 (Hsp70). A comparison of the identically colored elements of secondary structure reveals the overall similarity in structure despite the difference in biochemical activities. [Drawn from 1ATN.pdb and 1ATR.pdb.]

of homologous proteins for which at least one three-dimensional structure is known that regions and residues critical to protein function are more strongly conserved than are other residues. For example, each type of globin contains a bound heme group with an iron atom at its center. A histidine residue that interacts directly with this iron atom (residue 64 in human myoglobin) is conserved in all globins. After we have identified key residues or highly conserved sequences within a family of proteins, we can sometimes identify other family members even when the overall level of sequence similarity is below statistical significance. Thus it may be useful to generate a *sequence template*—a map of conserved residues that are structurally and functionally important and are characteristic of particular families of proteins, which makes it possible to recognize new family members that might be undetectable by other means. A variety of other methods for sequence classification that take advantage of known three-dimensional structures also are being developed. Still other methods are able to identify conserved residues within a family of homologous proteins, even without a known three-dimensional structure. These methods often use substitution matrices that differ at each position within a family of aligned sequences. Such methods can often detect quite distant evolutionary relationships.

Repeated motifs can be detected by aligning sequences with themselves

More than 10% of all proteins contain sets of two or more domains that are similar to one another. Sequence search methods can often detect internally repeated sequences that have been characterized in other proteins. Often, however, repeated units do not correspond to previously identified domains. In these cases, their presence can be detected by attempting to align a given sequence with itself. The statistical significance of such repeats can be tested by aligning the regions in question as if these regions were sequences from separate proteins. For the TATA-box-binding protein (Figure 6.17A), a key protein in controlling gene transcription (Section 29.2), such an alignment is highly significant: 30% of the amino acids are identical over 90 residues (Figure 6.17B). The

```
(A)   1  MTDQGLEGSN PVDLSKHPSG IVPTLQNIVS TVNLDCKLDL KAIALQARNA
     51  EYNPKRFAAV IMRIREPKTT ALIFASGKMV CTGAKSEDFS KMAARKYARI
    101  VQKLGFPAKF KDFKIQNIVG SCDVKFPIRL EGLAYSHAAF SSYEPELFPG
    151  LIYRMKVPKI VLLIFVSGKI VITGAKMRDE TYKAFENIYP VLSEFRKIQQ
```



```
(B)   1  MTDQGLEGSNPVDLSKHPS

     20  GIVPTLQNIVSTVNLDCKLDLKAIALQ—ARNAEYNPKRFAAVIMRIR
    110  FKDFKIQNIVGSCDVKFPIRLEGLAYSHAAFSSYEPELFPGLIYRMK

     66  EPKTTALIFASGKMVCTGAKSEDFSKMAARKYARIVQKLGFPAK
    157  VPKIVLLIFVSGKIVITGAKMRDETYKAFENIYPVLSEFRKIQQ
```

FIGURE 6.17 Sequence alignment of internal repeats. (A) The primary structure of the TATA-box-binding protein. (B) An alignment of the sequences of the two repeats of the TATA-box-binding protein. The amino-terminal repeat is shown in red and the carboxyl-terminal repeat in blue. (C) Structure of the TATA-box-binding protein. The amino-terminal domain is shown in red and the carboxyl-terminal domain in blue. [Drawn from 1VOK.pdb.]

estimated probability of such an alignment occurring by chance is 1 in 10^{13}. The determination of the three-dimensional structure of the TATA-box-binding protein confirmed the presence of repeated structures; the protein is formed of two nearly identical domains (Figure 6.17C). The evidence is convincing that the gene encoding this protein evolved by duplication of a gene encoding a single domain.

Convergent evolution illustrates common solutions to biochemical challenges

Thus far, we have been exploring proteins derived from common ancestors—that is, through *divergent evolution.* Other cases have been found of proteins that are structurally similar in important ways but are not descended from a common ancestor. How might two unrelated proteins come to resemble each other structurally? Two proteins evolving independently may have converged on similar structural features to perform a similar biochemical activity. Perhaps that structure was an especially effective solution to a biochemical problem that organisms face. The process by which very different evolutionary pathways lead to the same solution is called *convergent evolution.*

An example of convergent evolution is found among the serine proteases. These enzymes, to be considered in more detail in Chapter 9, cleave peptide bonds by hydrolysis. Figure 6.18 shows the structure of the active sites—that is, the sites on the proteins at which the hydrolysis reaction takes place—for two such enzymes, chymotrypsin and subtilisin. These active-site structures are remarkably similar. In each case, a serine residue, a histidine residue, and an aspartic acid residue are positioned in space in nearly identical arrangements. As we will see, this conserved spatial arrangement is critical for the activity of these enzymes and affords the same mechanistic solution to the problem of peptide hydrolysis. At first glance, this similarity might suggest that these proteins are homologous. However, striking differences in the overall structures of these proteins make an evolutionary relationship extremely unlikely (Figure 6.19). Whereas chymotrypsin consists almost entirely of β sheets, subtilisin contains extensive α-helical

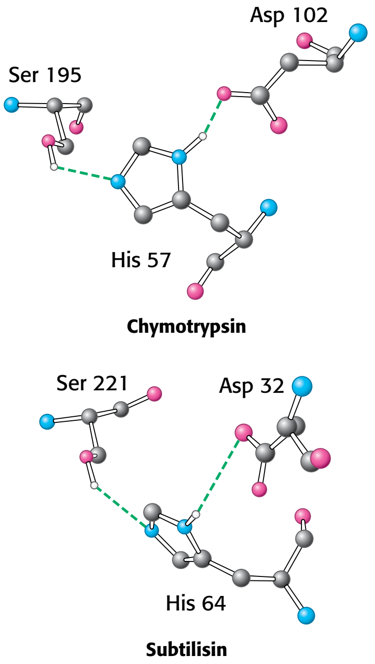


FIGURE 6.18 Convergent evolution of protease active sites. The relative positions of the three key residues shown are nearly identical in the active sites of the serine proteases chymotrypsin and subtilisin.

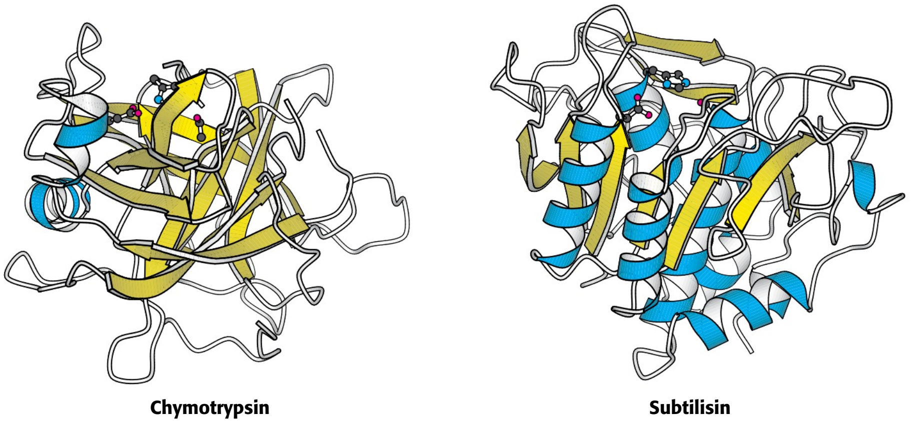

Chymotrypsin
Subtilisin

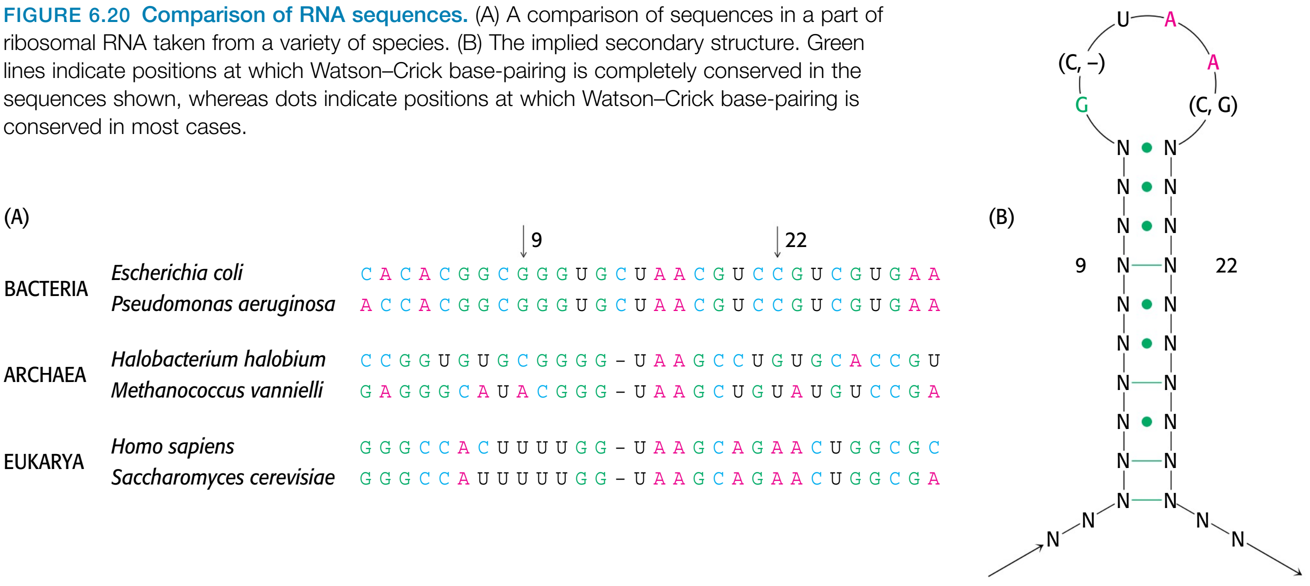
FIGURE 6.19 Structures of mammalian chymotrypsin and bacterial subtilisin. The overall structures are quite dissimilar, in stark contrast with the active sites, shown at the top of each structure. The β strands are shown in yellow and the α helices in blue. [Drawn from 1GCT.pdb. and 1SUP.pdb.]

structure. Moreover, the key serine, histidine, and aspartic acid residues do not occupy similar positions or even appear in the same order within the two sequences. It is extremely unlikely that two proteins evolving from a common ancestor could have retained similar active-site structures while other aspects of the structure changed so dramatically.

Comparison of RNA sequences can be a source of insight into RNA secondary structures

Homologous RNA sequences can be compared in a manner similar to that already described for protein sequences. Such comparisons can be a source of important insights into evolutionary relationships; in addition, they provide clues to the three-dimensional structure of the RNA itself. As noted in Chapter 4, single-stranded nucleic acid molecules fold back on themselves to form elaborate structures held together by Watson–Crick base-pairing and other interactions. In a family of sequences that form similar base-paired structures, base sequences may vary, but base-pairing ability is conserved. Consider, for example, a region from a large RNA molecule present in the ribosomes of all organisms (Figure 6.20). In the region shown, the *E. coli* sequence has a guanine (G) residue in position 9 and a cytosine (C) residue in position 22, whereas the human sequence has uracil (U) in position 9 and adenine (A) in position 22. Examination of the six sequences shown in Figure 6.20 reveals that the

FIGURE 6.20 Comparison of RNA sequences. (A) A comparison of sequences in a part of ribosomal RNA taken from a variety of species. (B) The implied secondary structure. Green lines indicate positions at which Watson–Crick base-pairing is completely conserved in the sequences shown, whereas dots indicate positions at which Watson–Crick base-pairing is conserved in most cases.


(A)
9
22
BACTERIA
Escherichia coli C A C A C G G C G G G U G C U A A C G U C C G U C G U G A A
Pseudomonas aeruginosa A C C A C G G C G G G U G C U A A C G U C C G U C G U G A A
ARCHAEA
Halobacterium halobium C C G G U G U G C G G G G – U A A G C C U G U G C A C C G U
Methanococcus vannielli G A G G G C A U A C G G G – U A A G C U G U A U G U C C G A
EUKARYA
Homo sapiens G G G C C A C U U U U G G – U A A G C A G A A C U G G C G C
Saccharomyces cerevisiae G G G C C A U U U U U G G – U A A G C A G A A C U G G C G A
(B)
U
A
(C, –)
A
G
(C, G)
N
9
22

bases in positions 9 and 22, as well as several of the neighboring positions, retain the ability to form Watson–Crick base pairs even though the identities of the bases in these positions vary. We can deduce that two segments with paired mutations that maintain base-pairing ability are likely to form a double helix. Where sequences are known for several homologous RNA molecules, this type of sequence analysis can often suggest complete secondary structures as well as some additional interactions. For this particular ribosomal RNA, the subsequent determination of its three-dimensional structure (Section 30.3) confirmed the predicted secondary structure.

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information

The observation that homology is often manifested as sequence similarity suggests that the evolutionary pathway relating the members of a family of proteins may be deduced by examination of sequence similarity. This approach is based on the notion that sequences that are more similar to one another have had less evolutionary time to diverge than have sequences that are less similar. This method can be illustrated by using the three globin sequences in Figures 6.11 and 6.13, as well as the sequence for the human hemoglobin β chain. These sequences can be aligned with the additional constraint that gaps, if present, should be at the same positions in all of the proteins. These aligned sequences can be used to construct an *evolutionary tree* in which the length of the branch connecting each pair of proteins is proportional to the number of amino acid differences between the sequences (Figure 6.21).

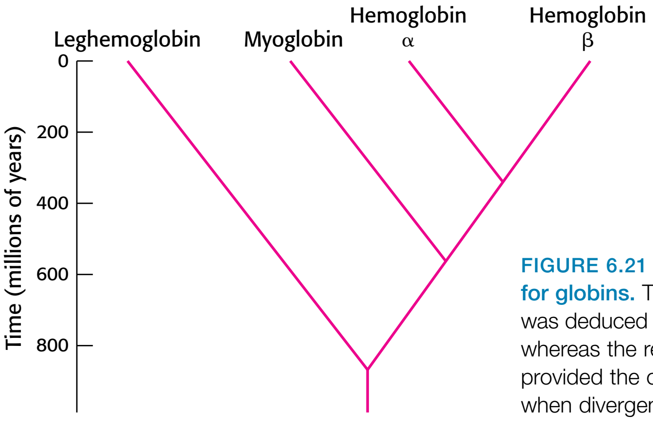


FIGURE 6.21 An evolutionary tree for globins. The branching structure was deduced by sequence comparison, whereas the results of fossil studies provided the overall time scale showing when divergence occurred.

Such comparisons reveal only the relative divergence times—for example, that myoglobin diverged from hemoglobin twice as long ago as the α chain diverged from the β chain. How can we estimate the approximate dates of gene duplications and other evolutionary events? Evolutionary trees can be calibrated by comparing the deduced branch points with divergence times determined from the fossil record. For example, the duplication leading to the two chains of hemoglobin appears to have occurred 350 million years ago. This estimate is supported by the observation that jawless fish such as the lamprey, which diverged from bony fish approximately 400 million years ago, contain hemoglobin built from a single type of subunit (Figure 6.22). These methods can be applied to both relatively modern and very ancient

FIGURE 6.22 The lamprey. A jawless fish whose ancestors diverged from bony fish approximately 400 million years ago, the lamprey has hemoglobin molecules that contain only a single type of polypeptide chain. [Brent P. Kent.]

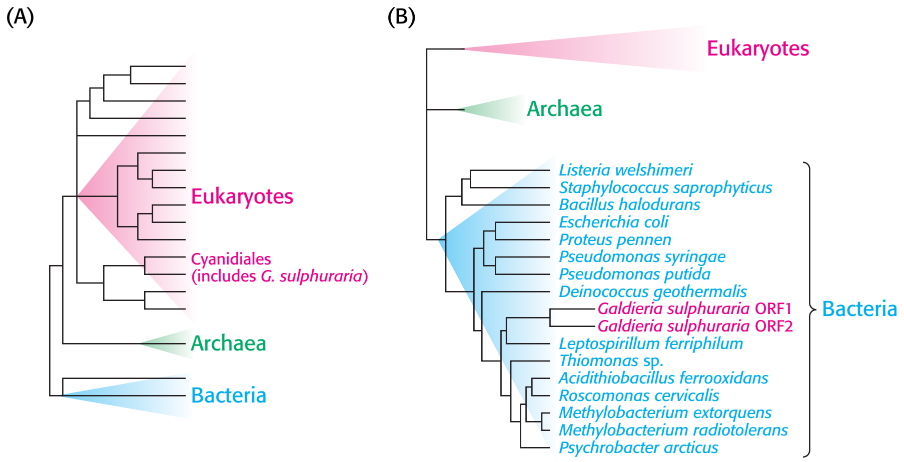

(A)
Eukaryotes
Cyanidiales
(includes G. sulphuraria)
Archaea
Bacteria
(B)
Eukaryotes
Archaea
Listeria welshimeri
Staphylococcus saprophyticus
Bacillus halodurans
Escherichia coli
Proteus pennen
Pseudomonas syringae
Pseudomonas putida
Deinococcus geothermalis
Galdieria sulphuraria ORF1
Galdieria sulphuraria ORF2
Leptospirillum ferriphilum
Thiomonas sp.
Acidithiobacillus ferrooxidans
Roscomonas cervicalis
Methylobacterium extorquens
Methylobacterium radiotolerans
Psychrobacter arcticus
Bacteria

FIGURE 6.23 Evidence of horizontal gene transfer. (A) The unicellular red alga *Galdieria sulphuraria* belongs to the order Cyanidiales, clearly within the eukaryotic branch of the evolutionary tree. (B) Within the completely sequenced *G. sulphuraria* genome, two ORFs encode proteins involved in transport of arsenate ions across membranes. Alignment of these ORFs against orthologs from a variety of species reveals that these pumps are most closely related to their bacterial counterparts, suggesting that a horizontal gene transfer event occurred during the evolution of this species. [(A) Information from Dr. Gerald Schönknecht; (B) Information from G. Schönknecht et al. *Science* 339:1207–1210, 2013, Fig. 3.]

molecules, such as the ribosomal RNAs that are found in all organisms. Indeed, such an RNA sequence analysis led to the realization that Archaea are a distinct group of organisms that diverged from Bacteria very early in evolutionary history.

Horizontal gene transfer events may explain unexpected branches of the evolutionary tree

Evolutionary trees that encompass orthologs of a particular protein across a range of species can lead to unexpected findings. For example, let us consider the unicellular red alga *Galdieria sulphuraria*, a remarkable eukaryote that can thrive in extreme environments, including at temperatures up to 56°C, at pH values between 0 and 4, and in the presence of high concentrations of toxic metals. *G. sulphuraria* belongs to the order Cyanidiales, clearly within the eukaryotic branch of the evolutionary tree (Figure 6.23A). However, the complete genome sequence of this organism revealed that nearly 5% of the *G. sulphuraria* ORFs encode proteins that are more closely related to bacterial or archaeal, not eukaryotic, orthologs. Furthermore, the proteins that exhibited these unexpected evolutionary relationships possess functions that are likely to confer a survival advantage in extreme environments, such as the removal of metal ions from inside the cell (Figure 6.23B). One likely explanation for these observations is *horizontal gene transfer*, or *the exchange of DNA between species* that provides a selective advantage to the recipient. Amongst prokaryotes, horizontal gene transfer is a well-characterized and important evolutionary mechanism. For example, as we shall discuss in Chapter 9, exchange of plasmid DNA between bacterial species likely facilitated the acquisition of restriction endonuclease activities. However, recent studies such as those on *G. sulphuraria*, made possible by the expansive growth of complete genome sequence information, suggest that horizontal gene transfer from prokaryotes to eukaryotes, between different domains of life, may also represent evolutionarily significant events.

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible

Two techniques of biochemistry have made it possible to examine the course of evolution more directly and not simply by inference. The polymerase chain reaction (Section 5.1) allows the direct examination of ancient DNA sequences, releasing us, at least in some cases, from the constraints of being able to examine existing genomes from living organisms only. Molecular evolution may be investigated through the use of *combinatorial chemistry,* the process of producing large populations of molecules en masse and selecting for a biochemical property. This exciting process provides a glimpse into the types of molecules that may have existed very early in evolution.

Ancient DNA can sometimes be amplified and sequenced

The tremendous chemical stability of DNA makes the molecule well suited to its role as the storage site of genetic information. So stable is the molecule that samples of DNA have survived for many thousands of years under appropriate conditions. With the development of PCR and advanced DNA-sequencing methods, such ancient DNA can be amplified and sequenced. This approach was first applied to mitochondrial DNA isolated from a Neanderthal fossil estimated at 38,000 years of age. Comparison of the complete Neanderthal mitochondrial sequence with those from *Homo sapiens* individuals revealed between 201 and 234 substitutions, considerably fewer than the approximately 1500 differences between human beings and chimpanzees over the same region.

Remarkably, the complete genome sequences of a Neanderthal and a closely related hominin known as a Denisovan have been obtained using DNA isolated from nearly 50,000-year-old fossils. Comparison of these sequences suggests that the common ancestor of modern human beings and Neanderthals lived approximately 570,000 years ago, while the common ancestor between Neanderthals and Denisovans lived nearly 380,000 years ago. An evolutionary tree constructed from these data revealed that the Neanderthal was not an intermediate between chimpanzees and human beings but, instead, was an evolutionary "dead end" that became extinct (Figure 6.24). Further analysis of these sequences has enabled researchers to determine the extent of interbreeding between these groups, elucidate the geographic history of these populations, and make assertions about additional ancestors whose DNA has not yet been sequenced.

A few earlier studies claimed to determine the sequences of far more ancient DNA such as that found in insects trapped in amber, but these studies appear to have been flawed. The source of these sequences turned out to be contaminating modern DNA. Successful sequencing of ancient DNA requires sufficient DNA for reliable amplification and the rigorous exclusion of all sources of contamination.

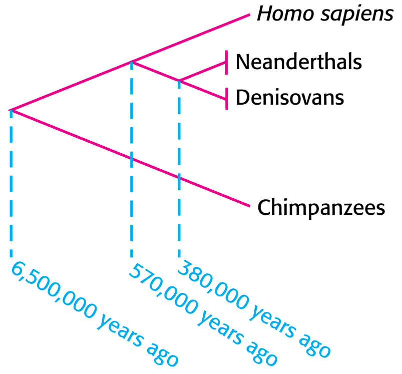


FIGURE 6.24 Placing Neanderthals and Denisovans on an evolutionary tree. Comparison of DNA sequences revealed that neither Neanderthals nor the Denisovans are on the line of direct descent leading to *Homo sapiens* but, instead, branched off earlier and then became extinct.

Molecular evolution can be examined experimentally

Evolution requires three processes: (1) the generation of a diverse population, (2) the selection of members based on some criterion of fitness, and (3) reproduction to enrich the population in these more-fit members. Nucleic acid molecules are capable of undergoing all three processes in vitro under appropriate conditions. The results of such studies enable us to glimpse how evolutionary processes might have generated catalytic activities and specific binding abilities—important biochemical functions in all living systems.

A diverse population of nucleic acid molecules can be synthesized in the laboratory by the process of combinatorial chemistry, which rapidly produces large populations of a particular type of molecule such as a nucleic acid. A population of molecules of a given size can be generated randomly so that many or all possible sequences are present in the mixture. When an initial population has been generated, it is subjected to a selection process that isolates specific molecules with desired binding or reactivity properties. Finally, molecules that have survived the selection process are replicated through the use of PCR; primers are directed toward specific sequences included at the ends of each member of the population. Errors that occur naturally in the course of the replication process introduce additional variation into the population in each "generation." Let us consider an application of this approach. Early in evolution, before the emergence of proteins, RNA molecules may have played all major roles in biological catalysis. To understand the properties of potential RNA catalysts, researchers have used the methods heretofore described to create an RNA molecule capable of binding adenosine triphosphate and related nucleotides. An initial population of RNA molecules 169 nucleotides long was created; 120 of the positions differed randomly, with equimolar mixtures of adenine, cytosine, guanine, and uracil. The initial synthetic pool that was used contained approximately 10^{14} RNA molecules. Note that this number is a very small fraction of the total possible pool of random 120-base sequences. From this pool, those molecules that bound to ATP, which had been immobilized on a column, were selected (Figure 6.25).

The collection of molecules that were bound well by the ATP affinity column was replicated by reverse transcription into DNA, amplification by PCR, and transcription back into RNA. The somewhat error-prone reverse transcriptase may have introduced additional mutations into the population

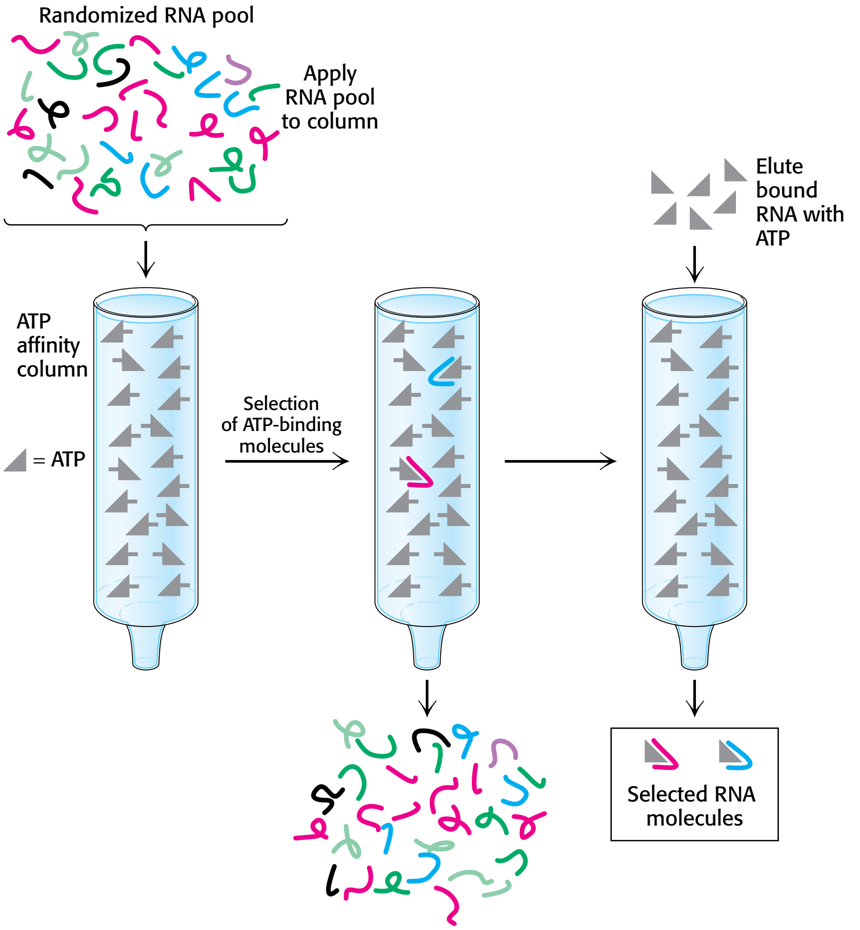


FIGURE 6.25 Evolution in the laboratory. A collection of RNA molecules of random sequences is synthesized by combinatorial chemistry. This collection is selected for the ability to bind ATP by passing the RNA through an ATP affinity column (Section 3.1). The ATP-binding RNA molecules are released from the column by washing with excess ATP and then replicated. The process of selection and replication is then repeated several times. The final RNA products with significant ATP-binding ability are isolated and characterized.

in each cycle. The new population was subjected to additional rounds of selection for ATP-binding activity. After eight generations, members of the selected population were characterized by sequencing. Seventeen different sequences were obtained, 16 of which could form the structure shown in Figure 6.26. Each of these molecules bound ATP with dissociation constants less than 50 μM.

The folded structure of the ATP-binding region from one of these RNAs was determined by nuclear magnetic resonance (NMR) methods (Section 3.6). As expected, this 40-nucleotide molecule is composed of two Watson–Crick base-paired helical regions separated by an 11-nucleotide loop (Figure 6.27A). This loop folds back on itself in an intricate way (Figure 6.27B) to form a deep pocket into which the adenine ring can fit (Figure 6.27C). Thus, a structure had evolved in vitro that was capable of a specific interaction.

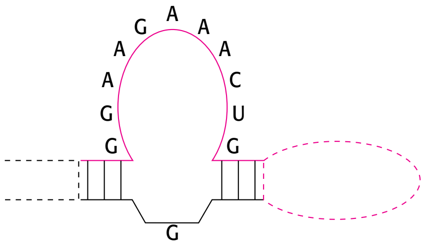


FIGURE 6.26 A conserved secondary structure. The secondary structure shown is common to RNA molecules selected for ATP binding.

Synthetic oligonucleotides that can specifically bind ligands, such as the ATP-binding RNA molecules described above, are referred to as *aptamers*. In addition to their role in understanding molecular evolution, aptamers have shown promise as versatile tools for biotechnology and medicine. They have been developed for diagnostic applications, serving as sensors for ligands ranging from small organic molecules, such as cocaine, to larger proteins, such as thrombin. Several aptamers are also in clinical trials as therapies for diseases ranging from leukemia to diabetes. Macugen (pegaptanib sodium), an aptamer which binds to and inhibits the protein vascular endothelial growth factor (VEGF), has been approved for the treatment of age-related macular degeneration.

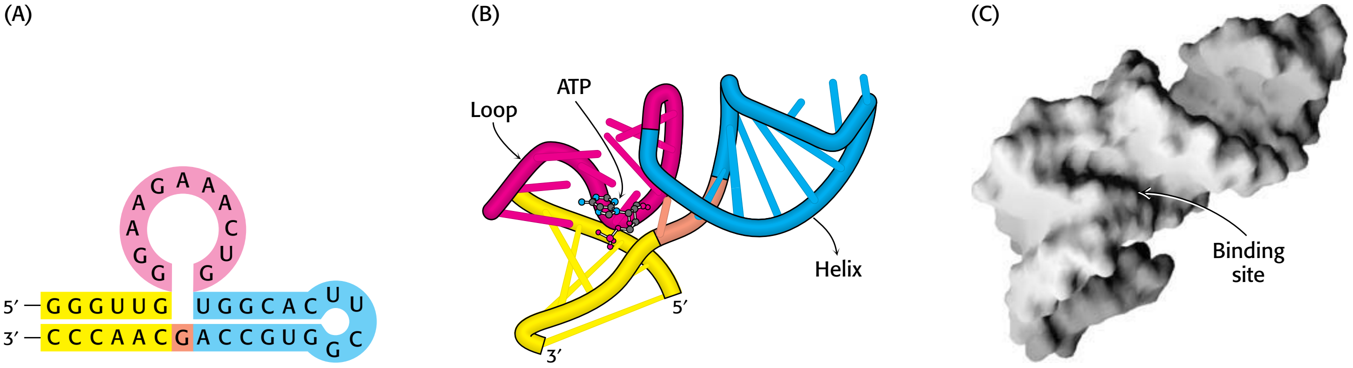


FIGURE 6.27 An evolved ATP-binding RNA molecule. (A) The Watson–Crick base-pairing pattern of an RNA molecule selected to bind adenosine nucleotides. (B) The NMR structure of this RNA molecule reveals the deep pocket into which the ATP molecule is bound. (C) In this surface representation, the ATP molecule has been removed to enable visualization of the pocket. [Drawn from 1RAW.pdb.]

SUMMARY

6.1 Homologs Are Descended from a Common Ancestor

Exploring evolution biochemically often means searching for homology between molecules, because homologous molecules, or homologs, evolved from a common ancestor. Paralogs are homologous molecules that are found in one species and have acquired different functions through evolutionary time. Orthologs are homologous molecules that are found in different species and have similar or identical functions.

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology

Protein and nucleic acid sequences are two of the primary languages of biochemistry. Sequence-alignment methods are the most powerful tools of the evolutionary detective. Sequences can be aligned to maximize their similarity, and the significance of these alignments can be judged by statistical tests. The detection of a statistically significant alignment between two sequences strongly suggests that two sequences are related by divergent evolution from a common ancestor. The use of substitution matrices makes the detection of more-distant evolutionary relationships possible. Any sequence can be used to probe sequence databases to identify related sequences present in the same organism or in other organisms.

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships

The evolutionary kinship between proteins may be even more strikingly evident in the conserved three-dimensional structures. The analysis of three-dimensional structure in combination with the analysis of especially conserved sequences has made it possible to determine evolutionary relationships that cannot be detected by other means. Sequence-comparison methods can also be used to detect imperfectly repeated sequences within a protein, indicative of linked similar domains.

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information

Evolutionary trees can be constructed with the assumption that the number of sequence differences corresponds to the time since the two sequences diverged. Construction of an evolutionary tree based on sequence comparisons revealed approximate times for the gene-duplication events separating myoglobin and hemoglobin as well as the α and β subunits of hemoglobin. Evolutionary trees based on sequences can be compared with those based on fossil records. Horizontal gene transfer events can manifest as unexpected branches on the evolutionary tree.

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible

The exploration of evolution can also be a laboratory science. In favorable cases, PCR amplification of well-preserved samples allows the determination of nucleotide sequences from extinct organisms. Sequences so determined can help authenticate parts of an evolutionary tree constructed by other means. Molecular evolutionary experiments performed in the test tube can examine how molecules such as ligand-binding RNA molecules might have been generated.

KEY TERMS

homolog (p. 170)
paralog (p. 170)
ortholog (p. 170)
sequence alignment (p. 171)
conservative substitution (p. 174)
nonconservative substitution (p. 174)
substitution matrix (p. 174)
BLAST search (p. 177)
sequence template (p. 180)
divergent evolution (p. 181)
convergent evolution (p. 181)
evolutionary tree (p. 183)
horizontal gene transfer (p. 184)
combinatorial chemistry (p. 185)
aptamers (p. 187)

PROBLEMS

1. *What's the score?* Using the identity-based scoring system (Section 6.2), calculate the score for the following alignment. Do you think the score is statistically significant?

```
(1) WYLGKITRMDAEVLLKKPTVRDGHFLVTQCESSPGEF-
(2) WYFGKITRRESERLLLNPENPRGTFLVRESETTKGAY-

    SISVRFGDSVQ-----HFKVLRDQNGKYYLWAVK-FN-
    CLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFS-

    SLNELVAYHRTASVSRTHTILLSDMNV
    SSLQQLVAYYSKHADGLCHRLTNV
```

2. *Sequence and structure.* A comparison of the aligned amino acid sequences of two proteins each consisting of 150 amino acids reveals them to be only 8% identical. However, their three-dimensional structures are very similar. Are these two proteins related evolutionarily? Explain.

3. *It depends on how you count.* Consider the following two sequence alignments:

```
(1) A-SNLFDIRLIG        (2) ASNLFDIRLI-G
    GSNDFYEVKIMD            GSNDFYEVKIMD
```

Which alignment has a higher score if the identity-based scoring system (Section 6.2) is used? Which alignment has a higher score if the Blosum-62 substitution matrix (Figure 6.9) is used?

4. *Discovering a new base pair.* Examine the ribosomal RNA sequences in Figure 6.20. In sequences that do not contain Watson–Crick base pairs, what base tends to be paired with G? Propose a structure for your new base pair.

5. *Overwhelmed by numbers.* Suppose that you wish to synthesize a pool of RNA molecules that contain all four bases at each of 40 positions. How much RNA must you have in grams if the pool is to have at least a single molecule of each sequence? The average molecular weight of a nucleotide is 330 g mol^{-1}.

6. *Form follows function.* The three-dimensional structure of biomolecules is more conserved evolutionarily than is sequence. Why?

7. *Shuffling.* Using the identity-based scoring system (Section 6.2), calculate the alignment score for the alignment of the following two short sequences:

(1) ASNFLDKAGK
(2) ATDYLEKAGK

Generate a shuffled version of sequence 2 by randomly reordering these 10 amino acids. Align your shuffled sequence with sequence 1 without allowing gaps, and calculate the alignment score between sequence 1 and your shuffled sequence.

8. *Interpreting the score.* Suppose that the sequences of two proteins each consisting of 200 amino acids are aligned and that the percentage of identical residues has been calculated. How would you interpret each of the following results in regard to the possible divergence of the two proteins from a common ancestor?

(a) 80%

(b) 50%

(c) 20%

(d) 10%.

9. *Particularly unique.* Consider the Blosum-62 matrix in Figure 6.9. Replacement of which three amino acids never yields a positive score? What features of these residues might contribute to this observation?

10. *A set of three.* The sequences of three proteins (A, B, and C) are compared with one another, yielding the following levels of identity:

	A	B	C
A	100%	65%	15%
B	65%	100%	55%
C	15%	55%	100%

Assume that the sequence matches are distributed uniformly along each aligned sequence pair. Would you expect protein A and protein C to have similar three-dimensional structures? Explain.

11. *RNA alignment.* Sequences of an RNA fragment from five species have been determined and aligned. Propose a likely secondary structure for these fragments.

(1) UUGGAGAUUCGGUAGAAUCUCCC
(2) GCCGGGAAUCGACAGAUUCCCCG
(3) CCCAAGUCCCGGCAGGGACUUAC
(4) CUCACCUGCCGAUAGGCAGGUCA
(5) AAUACCACCCGGUAGGGUGGUUC

12. *The more the merrier.* When RNA alignments are used to determine secondary structure, it is advantageous to have many sequences representing a wide variety of species. Why?

13. *To err is human.* You have discovered a mutant form of a thermostable DNA polymerase with significantly reduced fidelity in adding the appropriate nucleotide to the growing DNA strand, compared with wild-type DNA polymerase. How might this mutant be useful in the molecular-evolution experiments described in Section 6.5?

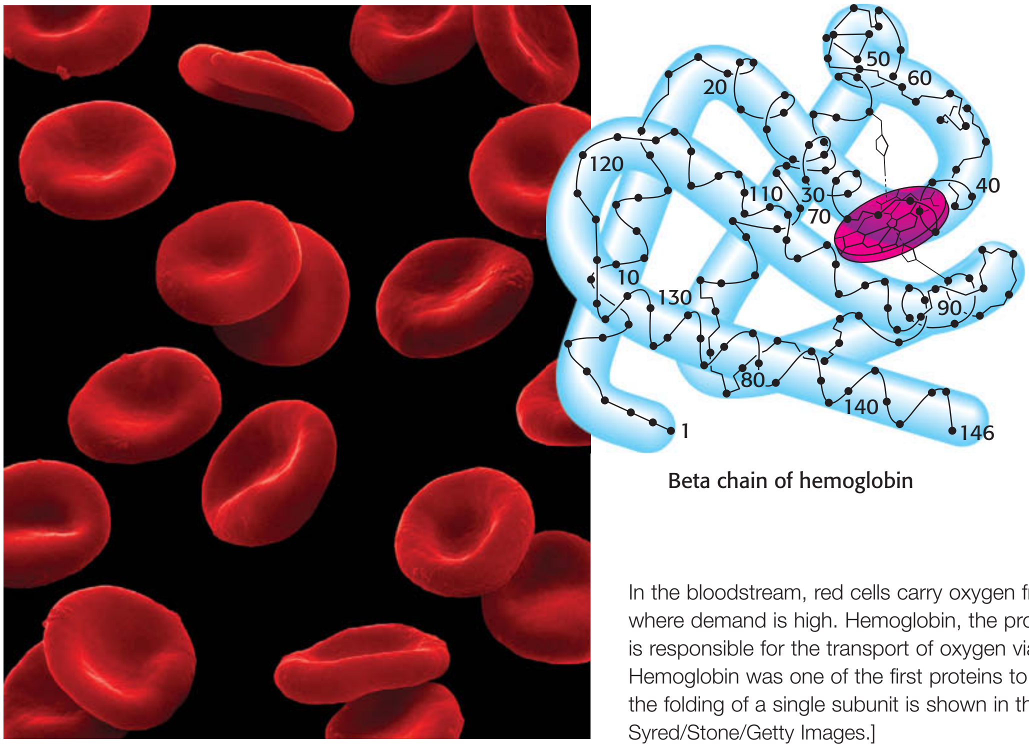
14. *Generation to generation.* When performing a molecular-evolution experiment, such as that described in Section 6.5, why is it important to repeat the selection and replication steps for several generations?

15. *BLAST away.* Using the National Center for Biotechnology Information Web site (www.ncbi.nlm.nih.gov), find the sequence of the enzyme triose phosphate isomerase from *E. coli* strain K-12. Use this sequence as the query for a protein–protein BLAST search. In the output, find the alignment with the sequence of triose phosphate isomerase from human beings (*Homo sapiens*). How many identities are observed in the alignment?

CHAPTER

7

Hemoglobin: Portrait of a Protein in Action



In the bloodstream, red cells carry oxygen from the lungs to the tissues, where demand is high. Hemoglobin, the protein that gives blood its red color, is responsible for the transport of oxygen via its four heme-bound subunits. Hemoglobin was one of the first proteins to have its structure determined; the folding of a single subunit is shown in this hand-drawn view. [Left, Andrew Syred/Stone/Getty Images.]

The transition from anaerobic to aerobic life was a major step in evolution because it uncovered a rich reservoir of energy. Fifteen times as much energy is extracted from glucose in the presence of oxygen than in its absence. For single-celled and other small organisms, oxygen can be absorbed into actively metabolizing cells directly from the air or surrounding water. Vertebrates evolved two principal mechanisms for supplying their cells with an adequate supply of oxygen. The first is a circulatory system that actively delivers oxygen to cells throughout the body. The second is the use of the oxygen-transport and oxygen-storage proteins, hemoglobin and myoglobin. Hemoglobin, which is contained in red blood cells, is a fascinating protein, efficiently carrying oxygen from the lungs to the tissues while also contributing to the transport of carbon dioxide and hydrogen ions back to the lungs. Myoglobin, located in muscle, facilitates the diffusion of oxygen through the cell for the generation of cellular energy and provides a reserve supply of oxygen available in time of need.

A comparison of myoglobin and hemoglobin illuminates some key aspects of protein structure and function. These two evolutionarily related proteins employ nearly identical structures for oxygen binding (Chapter 6). However, hemoglobin is a remarkably efficient oxygen carrier, able to use as much as 90% of its potential oxygen-carrying capacity effectively. Under similar conditions, myoglobin would be able to use only 7% of its potential capacity. What accounts

OUTLINE

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme

7.2 Hemoglobin Binds Oxygen Cooperatively

7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen: The Bohr Effect

7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease

for this dramatic difference? Myoglobin exists as a single polypeptide, whereas hemoglobin comprises four polypeptide chains. The four chains in hemoglobin bind oxygen *cooperatively,* meaning that the binding of oxygen to a site in one chain increases the likelihood that the remaining chains will bind oxygen. Furthermore, the oxygen-binding properties of hemoglobin are modulated by the binding of hydrogen ions and carbon dioxide in a manner that enhances oxygen-carrying capacity. Both cooperativity and the response to modulators are made possible by variations in the quaternary structure of hemoglobin when different combinations of molecules are bound.

Hemoglobin and myoglobin have played important roles in the history of biochemistry. They were the first proteins for which three-dimensional structures were determined by x-ray crystallography. Furthermore, the possibility that variations in protein sequence could lead to disease was first proposed and demonstrated for sickle-cell anemia, a blood disease caused by mutation of a single amino acid in one hemoglobin chain. Hemoglobin has been and continues to be a valuable source of knowledge and insight, both in itself and as a prototype for many other proteins that we will encounter throughout our study of biochemistry.

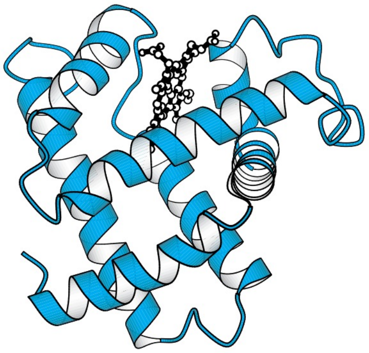
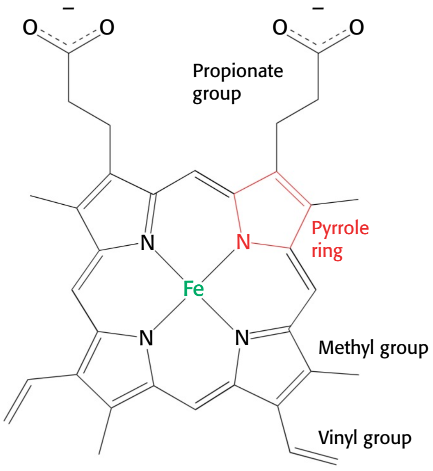


FIGURE 7.1 Structure of myoglobin. *Notice* that myoglobin consists of a single polypeptide chain, formed of α helices connected by turns, with one oxygen-binding site. [Drawn from 1MBD.pdb.]

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme

Sperm whale myoglobin was the first protein for which the three-dimensional structure was determined. X-ray crystallographic studies pioneered by John Kendrew revealed the structure of this protein in the 1950s (Figure 7.1). Myoglobin consists largely of α helices that are linked to one another by turns to form a globular structure.

Myoglobin can exist in an oxygen-free form called *deoxymyoglobin* or in a form with an oxygen molecule bound called *oxymyoglobin*. The ability of myoglobin and hemoglobin to bind oxygen depends on the presence of a *heme* molecule. As we shall discuss in Chapter 9, heme is one example of a prosthetic group, a molecule that binds tightly to a protein and is essential for its function.



Heme
(Fe-protoporphyrin IX)

The heme group gives muscle and blood their distinctive red color. It consists of an organic component and a central iron atom. The organic component, called *protoporphyrin,* is made up of four pyrrole rings linked by methine bridges to form a tetrapyrrole ring. Four methyl groups, two vinyl groups, and two propionate side chains are attached to the central tetrapyrrole.

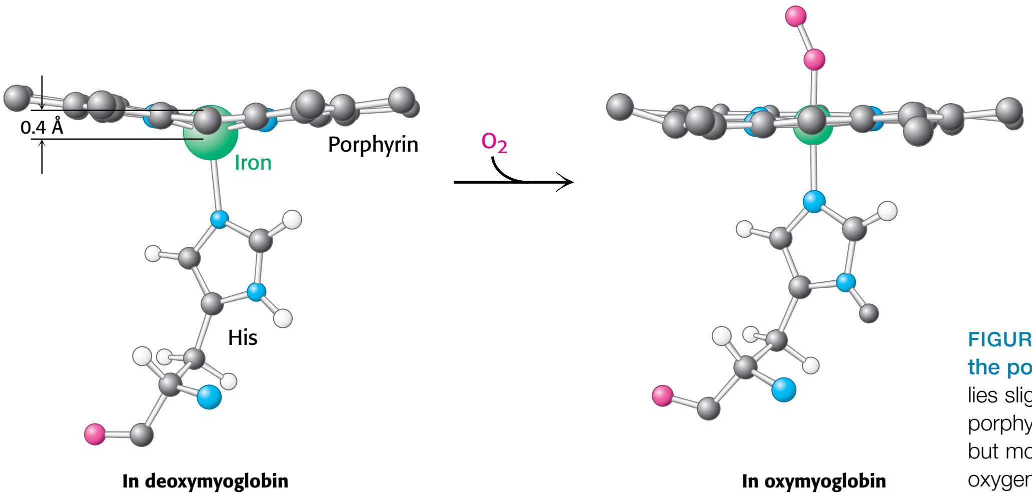


FIGURE 7.2 Oxygen binding changes the position of the iron ion. The iron ion lies slightly outside the plane of the porphyrin in deoxymyoglobin heme (left), but moves into the plane of the heme on oxygenation (right).

The iron atom lies in the center of the protoporphyrin, bonded to the four pyrrole nitrogen atoms. Although the heme-bound iron can be in either the ferrous (Fe^{2+}) or ferric (Fe^{3+}) oxidation state, only the Fe^{2+} state is capable of binding oxygen. The iron ion can form two additional bonds, one on each side of the heme plane. These binding sites are called the fifth and sixth coordination sites. In myoglobin, the fifth coordination site is occupied by the imidazole ring of a histidine residue from the protein. This histidine is referred to as the *proximal histidine*.

Oxygen binding occurs at the sixth coordination site. In deoxymyoglobin, this site remains unoccupied. The iron ion is slightly too large to fit into the well-defined hole within the porphyrin ring; it lies approximately 0.4 Å outside the porphyrin plane (Figure 7.2, left). Binding of the oxygen molecule at the sixth coordination site substantially rearranges the electrons within the iron so that the ion becomes effectively smaller, allowing it to move within the plane of the porphyrin (Figure 7.2, right). Remarkably, the structural changes that take place on oxygen binding were predicted by Linus Pauling, on the basis of magnetic measurements in 1936, nearly 25 years before the three-dimensional structures of myoglobin and hemoglobin were elucidated.

Changes in heme electronic structure upon oxygen binding are the basis for functional imaging studies

The change in electronic structure that occurs when the iron ion moves into the plane of the porphyrin is paralleled by alterations in the magnetic properties of hemoglobin; these changes are the basis for *functional magnetic resonance imaging* (fMRI), one of the most powerful methods for examining brain function. Nuclear magnetic resonance techniques detect signals that originate primarily from the protons in water molecules and are altered by the magnetic properties of hemoglobin. With the use of appropriate techniques, images can be generated that reveal differences in the relative amounts of deoxy- and oxyhemoglobin and thus the relative activity of various parts of the brain. When a specific part of the brain is active, blood vessels relax to allow more blood flow to that region. Thus, a more-active region of the brain will be richer in oxyhemoglobin.

These noninvasive methods identify areas of the brain that process sensory information. For example, subjects have been imaged while breathing air that either does or does not contain odorants. When odorants are present, fMRI detects an increase in the level of hemoglobin oxygenation (and, hence,

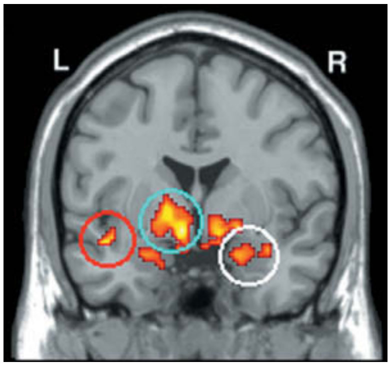


FIGURE 7.3 Functional magnetic resonance imaging of the brain. A functional magnetic resonance image reveals brain response to odorants. The light spots indicate regions of the brain activated by odorants. [From Osterbauer, R. et al., "Color of Scents: Chromatic Stimuli Modulate Odor Responses in the Human Brain." *Journal of Neurophysiology* 93 (2005) 3434–3441.]

of activity) in several regions of the brain (Figure 7.3). These regions are in the primary olfactory cortex, as well as in areas in which secondary processing of olfactory signals presumably takes place. Further analysis reveals the time course of activation of particular regions. Functional MRI shows tremendous potential for mapping regions and pathways engaged in processing sensory information obtained from all the senses. A seemingly incidental aspect of the biochemistry of hemoglobin has enabled observation of the brain in action.

The structure of myoglobin prevents the release of reactive oxygen species

Oxygen binding to iron in heme is accompanied by the partial transfer of an electron from the ferrous ion to oxygen. In many ways, the structure is best described as a complex between ferric ion (Fe^{3+}) and *superoxide anion* (O_2^-), as illustrated in Figure 7.4. It is crucial that oxygen, when it is released, leaves as dioxygen rather than superoxide, for two important reasons. First, superoxide and other species generated from it are *reactive oxygen species* that can be damaging to many biological materials. Second, release of superoxide would leave the iron ion in the ferric state. This species, termed *metmyoglobin*, does not bind oxygen. Thus, potential oxygen-storage capacity is lost. Features of myoglobin stabilize the oxygen complex such that superoxide is less likely to be released. In particular, the binding pocket of myoglobin includes an additional histidine residue (termed the *distal histidine*) that donates a hydrogen bond to the bound oxygen molecule (Figure 7.5). The superoxide character of the bound oxygen species strengthens this interaction. Thus, *the protein component of myoglobin controls the intrinsic reactivity*

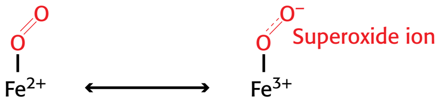


FIGURE 7.4 Iron–oxygen bonding. The interaction between iron and oxygen in myoglobin can be described as a combination of resonance structures, one with Fe^{2+} and dioxygen and another with Fe^{3+} and superoxide ion.

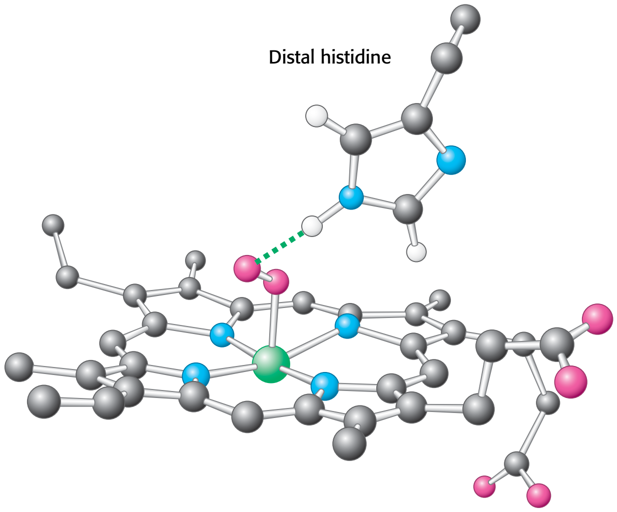


FIGURE 7.5 Stabilizing bound oxygen. A hydrogen bond (dotted green line) donated by the distal histidine residue to the bound oxygen molecule helps stabilize oxymyoglobin.

of heme, making it more suitable for reversible oxygen binding. The distal histidine may also impair access of carbon monoxide to the heme, which binds tightly to the heme iron with dire consequences (p. 201).

Human hemoglobin is an assembly of four myoglobin-like subunits

The three-dimensional structure of hemoglobin from horse heart was solved by Max Perutz shortly after the determination of the myoglobin structure. Since then, the structures of hemoglobins from other species including humans have been determined. Hemoglobin consists of four polypeptide chains, two identical *α chains* and two identical *β chains* (Figure 7.6). Each of the subunits consists of a set of α helices in the same arrangement as the α helices in myoglobin (see Figure 6.15 for a comparison of the structures). The recurring structure is called a *globin fold.* Consistent with this structural similarity, alignment of the amino acid sequences of the α and β chains of human hemoglobin with those of sperm whale myoglobin yields 25% and 24% identity, respectively, and good conservation of key residues such as the proximal and distal histidines. Thus, the α and β chains are related to each other and to myoglobin by divergent evolution (Section 6.2).

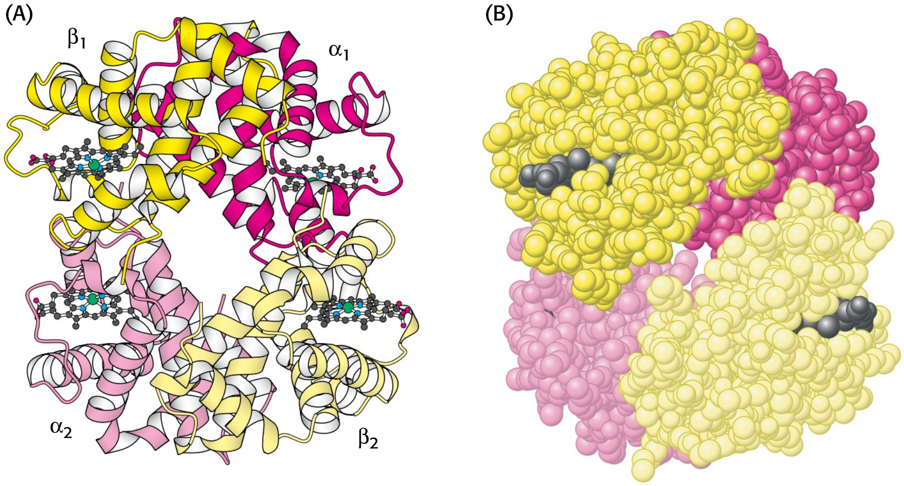


FIGURE 7.6 Quaternary structure of deoxyhemoglobin. Hemoglobin, which is composed of two α chains and two β chains, functions as a pair of αβ dimers. (A) A ribbon diagram. (B) A space-filling model. [Drawn from 1A3N.pdb.]

The hemoglobin tetramer, referred to as *hemoglobin A* (HbA), is best described as a pair of identical *αβ dimers* ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) that associate to form the tetramer. In deoxyhemoglobin, these αβ dimers are linked by an extensive interface, which includes the carboxyl terminus of each chain. The heme groups are well separated in the tetramer by iron–iron distances ranging from 24 to 40 Å.

7.2 Hemoglobin Binds Oxygen Cooperatively

We can determine the oxygen-binding properties of each of these proteins by observing its *oxygen-binding curve,* a plot of the *fractional saturation* versus the concentration of oxygen. The fractional saturation, Y, is defined as the fraction of possible binding sites that contain bound oxygen. The value of Y can range from 0 (all sites empty) to 1 (all sites filled). The concentration of oxygen is most conveniently measured by its *partial pressure,* pO_2.

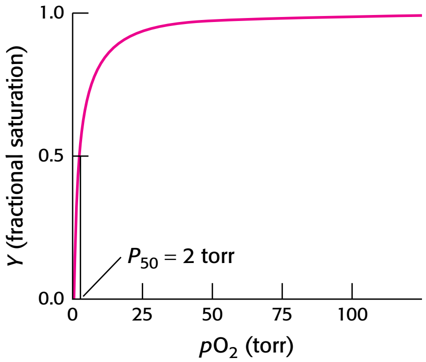


FIGURE 7.7 Oxygen binding by myoglobin. Half the myoglobin molecules have bound oxygen when the oxygen partial pressure is 2 torr.

Torr

A unit of pressure equal to that exerted by a column of mercury 1 mm high at 0°C and standard gravity (1 mm Hg). Named after Evangelista Torricelli (1608–1647), inventor of the mercury barometer.

For myoglobin, a binding curve indicating a simple chemical equilibrium is observed (Figure 7.7). Notice that the curve rises sharply as pO_2 increases and then levels off. Half-saturation of the binding sites, referred to as P_{50} (for 50% saturated), is at the relatively low value of 2 torr (mm Hg), indicating that oxygen binds with high affinity to myoglobin.

In contrast, the oxygen-binding curve for hemoglobin in red blood cells shows some remarkable features (Figure 7.8). It does not look like a simple binding curve such as that for myoglobin; instead, it resembles an "S." Such curves are referred to as *sigmoid* because of their S-like shape. In addition, oxygen binding for hemoglobin (P_{50} = 26 torr) is significantly weaker than that for myoglobin. Note that this binding curve is derived from hemoglobin in red blood cells.

A sigmoid binding curve indicates that a protein exhibits a special binding behavior. For hemoglobin, this shape suggests that the binding of oxygen at one site within the hemoglobin tetramer increases the likelihood that oxygen binds at the remaining unoccupied sites. Conversely, the unloading of oxygen at one heme facilitates the unloading of oxygen at the others. This sort of binding behavior is referred to as *cooperative*, because the binding reactions at individual sites in each hemoglobin molecule are not independent of one another. We will return to the mechanism of this cooperativity shortly.

What is the physiological significance of the cooperative binding of oxygen by hemoglobin? Oxygen must be transported in the blood from the lungs, where the partial pressure of oxygen is relatively high (approximately 100 torr), to the actively metabolizing tissues, where the partial pressure of oxygen is much lower (typically, 20 torr). Let us consider how the cooperative behavior indicated by the sigmoid curve leads to efficient oxygen transport (Figure 7.9). In the lungs, hemoglobin becomes nearly

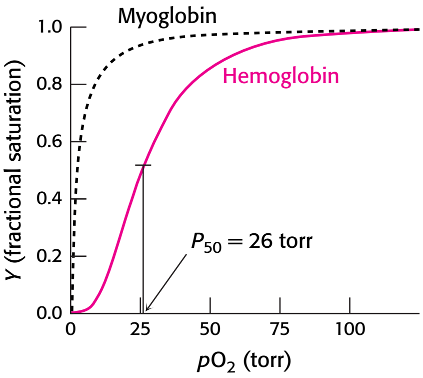


FIGURE 7.8 Oxygen binding by hemoglobin. This curve, obtained for hemoglobin in red blood cells, is shaped somewhat like an "S," indicating that distinct, but interacting, oxygen-binding sites are present in each hemoglobin molecule. Half-saturation for hemoglobin is 26 torr. For comparison, the binding curve for myoglobin is shown as a dashed black curve.

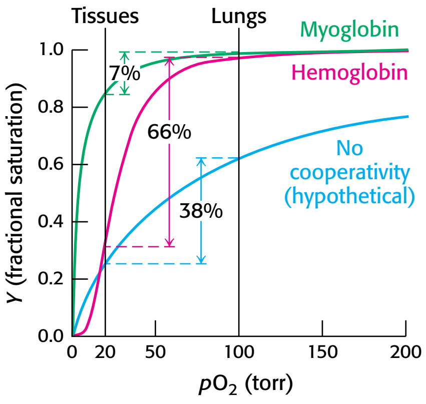


FIGURE 7.9 Cooperativity enhances oxygen delivery by hemoglobin. Because of cooperativity between O_2 binding sites, hemoglobin delivers more O_2 to actively metabolizing tissues than would myoglobin or any noncooperative protein, even one with optimal O_2 affinity.

saturated with oxygen such that 98% of the oxygen-binding sites are occupied. When hemoglobin moves to the tissues and releases O_2, the saturation level drops to 32%. Thus, a total of $98 - 32 = 66\%$ of the potential oxygen-binding sites contribute to oxygen transport. The cooperative release of oxygen favors a more-complete unloading of oxygen in the tissues. If myoglobin were employed for oxygen transport, it would be 98% saturated in the lungs, but would remain 91% saturated in the tissues, and so only $98 - 91 = 7\%$ of the sites would contribute to oxygen transport; myoglobin binds oxygen too tightly to be useful in oxygen transport. Nature might have solved this problem by weakening the affinity of myoglobin for oxygen to maximize the difference in saturation between 20 and 100 torr. However, for such a protein, the most oxygen that could be transported from a region in which pO_2 is 100 torr to one in which it is 20 torr is $63 - 25 = 38\%$, as indicated by the blue curve in Figure 7.9. Thus, the cooperative binding and release of oxygen by hemoglobin enables it to deliver nearly 10 times as much oxygen as could be delivered by myoglobin and more than 1.7 times as much as could be delivered by any noncooperative protein.

Closer examination of oxygen concentrations in tissues at rest and during exercise underscores the effectiveness of hemoglobin as an oxygen carrier (Figure 7.10). Under resting conditions, the oxygen concentration in muscle is approximately 40 torr, but during exercise the concentration is reduced to 20 torr. In the decrease from 100 torr in the lungs to 40 torr in resting muscle, the oxygen saturation of hemoglobin is reduced from 98% to 77%, and so $98 - 77 = 21\%$ of the oxygen is released over a drop of 60 torr. In a decrease from 40 torr to 20 torr, the oxygen saturation is reduced from 77% to 32%, corresponding to an oxygen release of 45% over a drop of 20 torr. Thus, because the change in oxygen concentration from rest to exercise corresponds to the steepest part of the oxygen-binding curve, oxygen is effectively delivered to tissues where it is most needed. In Section 7.3, we shall examine other properties of hemoglobin that enhance its physiological responsiveness.

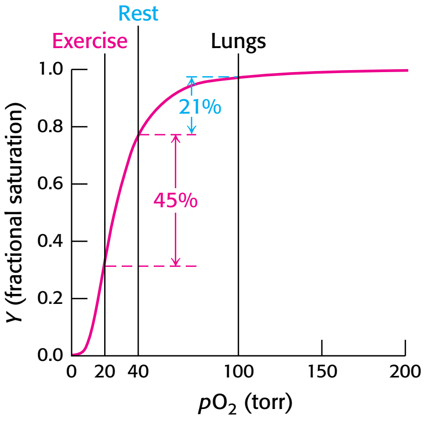


FIGURE 7.10 Responding to exercise. The drop in oxygen concentration from 40 torr in resting tissues to 20 torr in exercising tissues corresponds to the steepest part of the observed oxygen-binding curve. As shown here, hemoglobin is very effective in providing oxygen to exercising tissues.

Oxygen binding markedly changes the quaternary structure of hemoglobin

The cooperative binding of oxygen by hemoglobin requires that the binding of oxygen at one site in the hemoglobin tetramer influence the oxygen-binding properties at the other sites. Given the large separation between the iron sites, direct interactions are not possible. Thus, indirect mechanisms for coupling the sites must be at work. These mechanisms are intimately related to the quaternary structure of hemoglobin.

Hemoglobin undergoes substantial changes in quaternary structure on oxygen binding: the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers rotate approximately 15 degrees with respect to one another (Figure 7.11). The dimers themselves are relatively unchanged, although there are localized conformational shifts. Thus, the interface between the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers is most affected by this structural transition. In particular, the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers are freer to move with respect to one another in the oxygenated state than they are in the deoxygenated state.

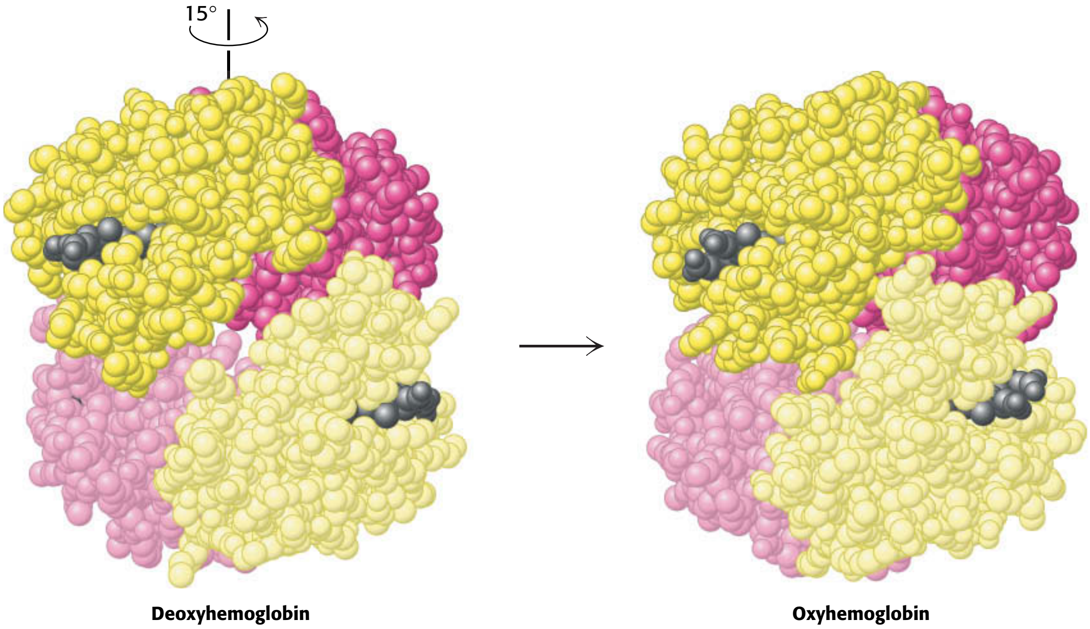


FIGURE 7.11 Quaternary structural changes on oxygen binding by hemoglobin. *Notice* that, on oxygenation, one αβ dimer shifts with respect to the other by a rotation of 15 degrees. [Drawn from 1A3N.pdb and 1LFQ.pdb.]

The quaternary structure observed in the deoxy form of hemoglobin, *deoxyhemoglobin*, is often referred to as the *T* (for tense) *state* because it is quite constrained by subunit–subunit interactions. The quaternary structure of the fully oxygenated form, *oxyhemoglobin*, is referred to as the *R* (for relaxed) *state*. In light of the observation that the R form of hemoglobin is less constrained, the tense and relaxed designations seem particularly apt. Importantly, in the R state, the oxygen-binding sites are free of strain and are capable of binding oxygen with higher affinity than are the sites in the T state. *By triggering the shift of the hemoglobin tetramer from the T state to the R state, the binding of oxygen to one site increases the binding affinity of other sites.*

Hemoglobin cooperativity can be potentially explained by several models

Two limiting models have been developed to explain the cooperative binding of ligands to a multisubunit assembly such as hemoglobin. In the *concerted model,* also known as the *MWC model* after Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux, who first proposed it, the overall assembly can exist only in two forms: the T state and the R state. The binding of ligands simply shifts the equilibrium between these two states (Figure 7.12).

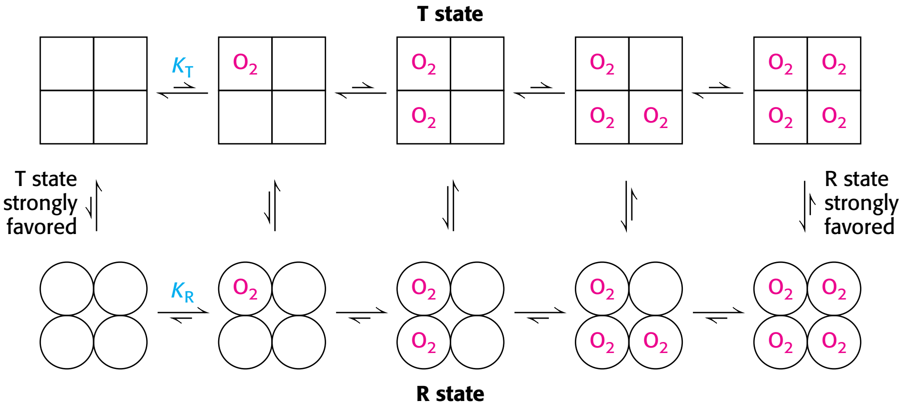


FIGURE 7.12 Concerted model. All molecules exist either in the T state or in the R state. At each level of oxygen loading, an equilibrium exists between the T and R states. The equilibrium shifts from strongly favoring the T state with no oxygen bound to strongly favoring the R state when the molecule is fully loaded with oxygen. The R state has a greater affinity for oxygen than does the T state.

Thus, as a hemoglobin tetramer binds each oxygen molecule, the probability that the tetramer is in the R state increases. Deoxyhemoglobin tetramers are almost exclusively in the T state. However, the binding of oxygen to one site in the molecule shifts the equilibrium toward the R state. If a molecule assumes the R quaternary structure, the oxygen affinity of its sites increases. Additional oxygen molecules are now more likely to bind to the three unoccupied sites. Thus, the binding curve for hemoglobin can be seen as a combination of the binding curves that would be observed if all molecules remained in the T state or if all of the molecules were in the R state. As oxygen molecules bind, the hemoglobin tetramers convert from the T state into the R state, yielding the sigmoid binding curve so important for efficient oxygen transport (Figure 7.13).

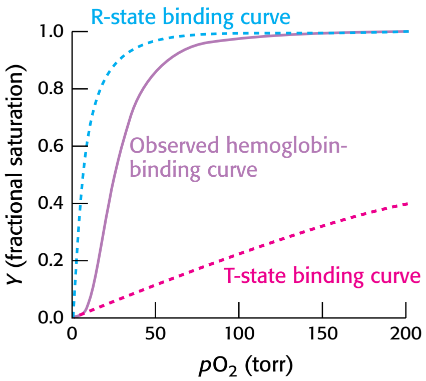


FIGURE 7.13 T-to-R transition. The binding curve for hemoglobin can be seen as a combination of the binding curves that would be observed if all molecules remained in the T state or if all of the molecules were in the R state. The sigmoidal curve is observed because molecules convert from the T state into the R state as oxygen molecules bind.

In the concerted model, each tetramer can exist in only two states, the T state and the R state. In an alternative model, the *sequential model,* the binding of a ligand to one site in an assembly increases the binding affinity of neighboring sites without inducing a full conversion from the T into the R state (Figure 7.14).

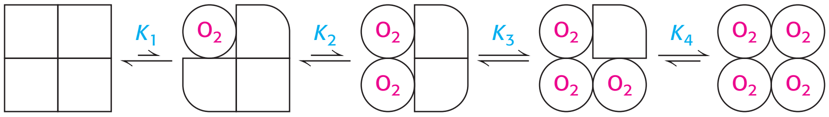


FIGURE 7.14 Sequential model. The binding of a ligand changes the conformation of the subunit to which it binds. This conformational change induces changes in neighboring subunits that increase their affinity for the ligand.

Is the cooperative binding of oxygen by hemoglobin better described by the concerted or the sequential model? Neither model in its pure form fully accounts for the behavior of hemoglobin. Instead, a combined model is required. Hemoglobin behavior is concerted in that the tetramer with three sites occupied by oxygen is almost always in the quaternary structure associated with the R state. The remaining open binding site has an affinity for oxygen more than 20-fold greater than that of fully deoxygenated hemoglobin binding its first oxygen. However, the behavior is not fully concerted, because hemoglobin with oxygen bound to only one of four sites remains primarily in the T-state quaternary structure. Yet, this molecule binds oxygen three times as strongly as does fully deoxygenated hemoglobin, an observation consistent only with a sequential model. These results highlight the fact that the concerted and sequential models represent idealized limiting cases, which real systems may approach but rarely attain.

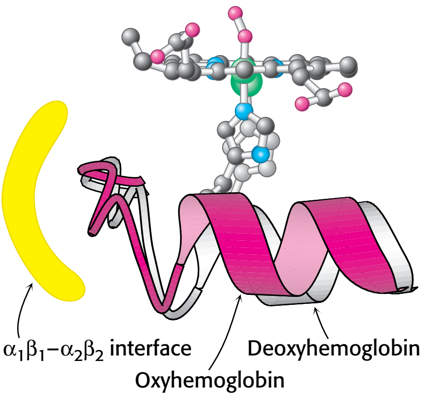


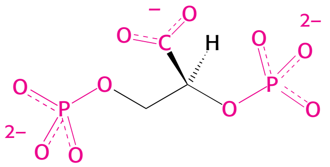
FIGURE 7.15 Conformational changes in hemoglobin. The movement of the iron ion on oxygenation brings the iron-associated histidine residue toward the porphyrin ring. The associated movement of the histidine-containing α helix alters the interface between the αβ dimers, instigating other structural changes. For comparison, the deoxyhemoglobin structure is shown in gray behind the oxyhemoglobin structure in red.

Structural changes at the heme groups are transmitted to the $\alpha_1\beta_1$–$\alpha_2\beta_2$ interface

We now examine how oxygen binding at one site is able to shift the equilibrium between the T and R states of the entire hemoglobin tetramer. As in myoglobin, oxygen binding causes each iron atom in hemoglobin to move from outside the plane of the porphyrin into the plane. When the iron atom moves, the proximal histidine residue moves with it. This histidine residue is part of an α helix, which also moves (Figure 7.15). The carboxyl terminal end of this α helix lies in the interface between the two αβ dimers. The change in position of the carboxyl terminal end of the helix favors the T-to-R transition. Consequently, *the structural transition at the iron ion in one subunit is directly transmitted to the other subunits.* The rearrangement of the dimer interface provides a pathway for communication between subunits, enabling the cooperative binding of oxygen.

2,3-Bisphosphoglycerate in red cells is crucial in determining the oxygen affinity of hemoglobin

For hemoglobin to function efficiently, the T state must remain stable until the binding of sufficient oxygen has converted it into the R state. In fact, however, the T state of hemoglobin is highly unstable, pushing the equilibrium so far toward the R state that little oxygen would be released in physiological conditions. Thus, an additional mechanism is needed to properly stabilize the T state. This mechanism was discovered by comparing the oxygen-binding properties of hemoglobin in red blood cells with fully purified hemoglobin (Figure 7.16). Pure hemoglobin binds oxygen much more tightly than does hemoglobin in red blood cells. This dramatic difference is due to the presence within these cells of *2,3-bisphosphoglycerate* (2,3-BPG; also known as 2,3-diphosphoglycerate or 2,3-DPG).



2,3-Bisphosphoglycerate (2,3-BPG)

This highly anionic compound is present in red blood cells at approximately the same concentration as that of hemoglobin (~2 mM). Without 2,3-BPG, hemoglobin would be an extremely inefficient oxygen transporter, releasing only 8% of its cargo in the tissues.

How does 2,3-BPG lower the oxygen affinity of hemoglobin so significantly? Examination of the crystal structure of deoxyhemoglobin in the presence of 2,3-BPG reveals that a single molecule of 2,3-BPG binds in the center of the tetramer, in a pocket present only in the T form (Figure 7.17). On T-to-R transition, this pocket collapses and 2,3-BPG is released. Thus, in order for the structural transition from T to R to take place, the bonds between hemoglobin and 2,3-BPG must be broken. In the presence of 2,3-BPG, more oxygen-binding sites within the hemoglobin tetramer must be occupied in order to induce the T-to-R transition, and so hemoglobin remains in the lower-affinity T state until higher oxygen concentrations are reached. This mechanism of regulation is remarkable because 2,3-BPG does not in any way resemble oxygen, the molecule on which hemoglobin carries out its primary function. 2,3-BPG is referred to as an *allosteric effector* (from

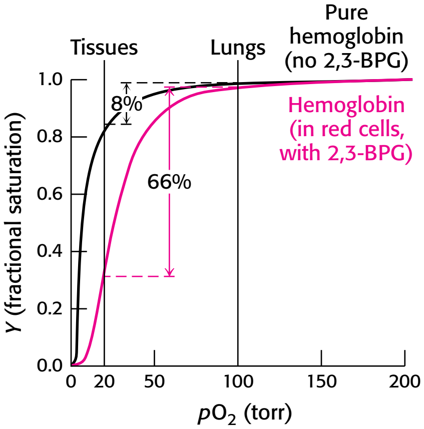


FIGURE 7.16 Oxygen binding by pure hemoglobin compared with hemoglobin in red blood cells. Pure hemoglobin binds oxygen more tightly than does hemoglobin in red blood cells. This difference is due to the presence of 2,3-bisphosphoglycerate (2,3-BPG) in red blood cells.

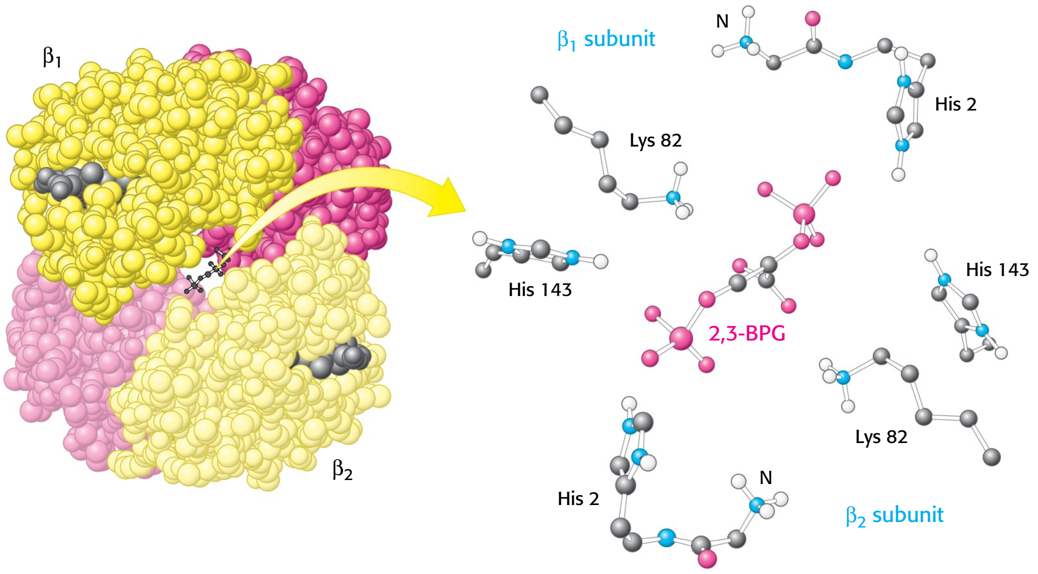


FIGURE 7.17 Mode of binding of 2,3-BPG to human deoxyhemoglobin. 2,3-Bisphosphoglycerate binds to the central cavity of deoxyhemoglobin (left). There, it interacts with three positively charged groups on each β chain (right). [Drawn from 1B86.pdb.]

the Greek *allos,* "other," and *stereos,* "structure"). Regulation by a molecule structurally unrelated to oxygen is possible because the allosteric effector binds to a site that is completely distinct from that for oxygen. We will encounter allosteric effects again when we consider enzyme regulation in Chapter 10.

The binding of 2,3-BPG to hemoglobin has other crucial physiological consequences. The globin gene expressed by human fetuses differs from that expressed by adults; *fetal hemoglobin* tetramers include two α chains and two γ chains. The γ chain, a result of gene duplication, is 72% identical in amino acid sequence with the β chain. One noteworthy change is the substitution of a serine residue for His 143 in the β chain, part of the 2,3-BPG-binding site. This change removes two positive charges from the 2,3-BPG-binding site (one from each chain) and reduces the affinity of 2,3-BPG for fetal hemoglobin. Consequently, the oxygen-binding affinity of fetal hemoglobin is higher than that of maternal (adult) hemoglobin (Figure 7.18). This difference in oxygen affinity allows oxygen to be effectively transferred from maternal to fetal red blood cells. We have here an example in which gene duplication and specialization produced a ready solution to a biological challenge—in this case, the transport of oxygen from mother to fetus.

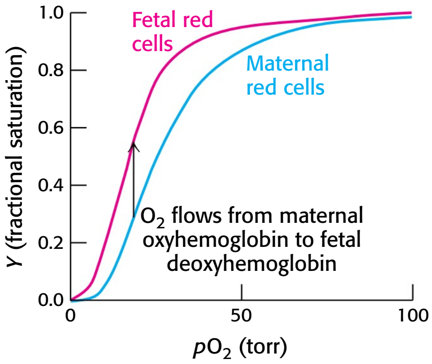


FIGURE 7.18 Oxygen affinity of fetal red blood cells. Fetal red blood cells have a higher oxygen affinity than do maternal red blood cells because fetal hemoglobin does not bind 2,3-BPG as well as maternal hemoglobin does.

Carbon monoxide can disrupt oxygen transport by hemoglobin

Carbon monoxide (CO) is a colorless, odorless gas that binds to hemoglobin at the same site as oxygen, forming a complex termed *carboxyhemoglobin*. Formation of carboxyhemoglobin exerts devastating effects on normal oxygen transport in two ways. First, carbon monoxide binds to hemoglobin about 200-fold more tightly than does oxygen. Even at low partial pressures in the blood, carbon monoxide will displace oxygen from hemoglobin, preventing its delivery. Second, carbon monoxide bound to one site in hemoglobin will shift the oxygen saturation curve of the remaining sites to the left, forcing the tetramer into the R state. This results in an increased affinity for oxygen, preventing its dissociation at tissues.

Exposure to carbon monoxide—from gas appliances and running automobiles, for example—can cause carbon monoxide poisoning, in which patients exhibit nausea, vomiting, lethargy, weakness, and disorientation.

One treatment for carbon monoxide poisoning is administration of 100% oxygen, often at pressures greater than atmospheric pressure (this treatment is referred to as *hyperbaric oxygen therapy*). With this therapy, the partial pressure of oxygen in the blood becomes sufficiently high to increase substantially the displacement of carbon monoxide from hemoglobin. Exposure to high concentrations of carbon monoxide, however, can be rapidly fatal: in the United States, about 2,500 people die each year from carbon monoxide poisoning, about 500 of them from accidental exposures and nearly 2,000 by suicide.

7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen: The Bohr Effect

We have seen how cooperative release of oxygen from hemoglobin helps deliver oxygen to where it is most needed: tissues exhibiting low oxygen partial pressures. This ability is enhanced by the facility of hemoglobin to respond to other cues in its physiological environment that signal the need for oxygen. Rapidly metabolizing tissues, such as contracting muscle, generate large amounts of hydrogen ions and carbon dioxide (Chapter 16). To release oxygen where the need is greatest, hemoglobin has evolved to respond to higher levels of these substances. Like 2,3-BPG, hydrogen ions and carbon dioxide are allosteric effectors of hemoglobin that bind to sites on the molecule that are distinct from the oxygen-binding sites. The regulation of oxygen binding by hydrogen ions and carbon dioxide is called the *Bohr effect* after Christian Bohr, who described this phenomenon in 1904.

The oxygen affinity of hemoglobin decreases as pH decreases from a value of 7.4 (Figure 7.19). Consequently, as hemoglobin moves into a region of lower pH, its tendency to release oxygen increases. For example, transport from the lungs, with pH 7.4 and an oxygen partial pressure of 100 torr, to active muscle, with a pH of 7.2 and an oxygen partial pressure of 20 torr, results in a release of oxygen amounting to 77% of total carrying capacity. Only 66% of the oxygen would be released in the absence of any change in pH. Structural and chemical studies have revealed much about the chemical basis of the Bohr effect. Several chemical groups within the hemoglobin tetramer are important for sensing changes in pH; all of these have pK_a values near pH 7. Consider histidine β146, the residue at the C terminus of the β chain. In deoxyhemoglobin, the terminal carboxylate group of β146 forms a ionic bond, also called a salt bridge, with a lysine residue in the α subunit of the other αβ dimer. This interaction locks the side chain of histidine β146 in a position from which it can participate in a salt bridge with

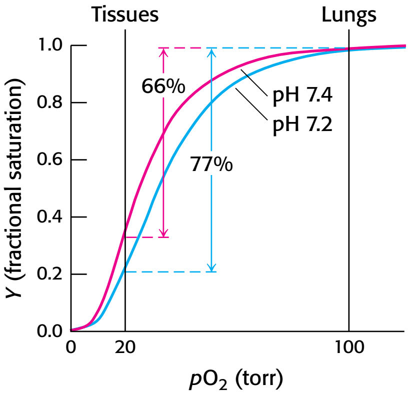


FIGURE 7.19 Effect of pH on the oxygen affinity of hemoglobin. Lowering the pH from 7.4 (red curve) to 7.2 (blue curve) results in the release of O_2 from oxyhemoglobin.

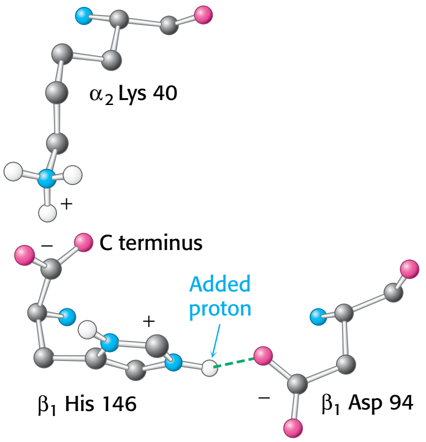


FIGURE 7.20 Chemical basis of the Bohr effect. In deoxyhemoglobin, three amino acid residues form two salt bridges that stabilize the T quaternary structure. The formation of one of the salt bridges depends on the presence of an added proton on histidine β146. The proximity of the negative charge on aspartate β94 in deoxyhemoglobin favors protonation of this histidine. *Notice* that the salt bridge between histidine β146 and aspartate β94 is stabilized by a hydrogen bond (green dashed line).

negatively charged aspartate β94 in the same chain, provided that the imidazole group of the histidine residue is protonated (Figure 7.20).

In addition to His β146, the α-amino groups at the amino termini of the α chain and the side chain of histidine α122 also participate in salt bridges in the T state. *The formation of these salt bridges stabilizes the T state, leading to a greater tendency for oxygen to be released.* For example, at high pH, the side chain of histidine β146 is not protonated and the salt bridge does not form. As the pH drops, however, the side chain of histidine β146 becomes protonated, the salt bridge with aspartate β94 forms, and the T state is stabilized.

Carbon dioxide, a neutral species, passes through the red-blood-cell membrane into the cell. This transport is also facilitated by membrane transporters, including proteins associated with Rh blood types. Carbon dioxide stimulates oxygen release by two mechanisms. First, the presence of high concentrations of carbon dioxide leads to a drop in pH within the red blood cell (Figure 7.21). Carbon dioxide reacts with water to form carbonic acid, H_2CO_3. This reaction is accelerated by *carbonic anhydrase,* an enzyme abundant in red blood cells that will be considered extensively in Chapter 9. H_2CO_3 is a moderately strong acid with a pK_a of 3.5. Thus, once formed, carbonic acid dissociates to form bicarbonate ion, HCO_3^-, and H^+, resulting in a drop in pH that stabilizes the T state by the mechanism discussed previously.

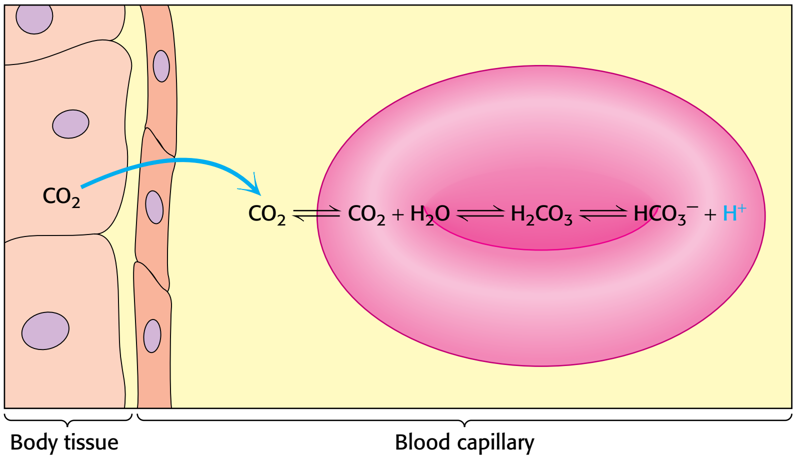

CO_2
$CO_2 \rightleftharpoons CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+$
Body tissue
Blood capillary

FIGURE 7.21 Carbon dioxide and pH. Carbon dioxide in the tissues diffuses into red blood cells. Inside a red blood cell, carbon dioxide reacts with water to form carbonic acid, in a reaction catalyzed by the enzyme carbonic anhydrase. Carbonic acid dissociates to form HCO_3^- and H^+, resulting in a drop in pH inside the red cell.

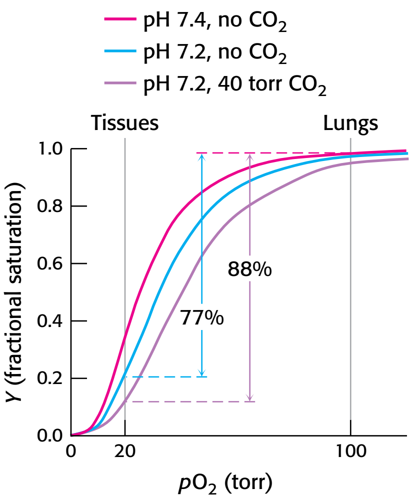
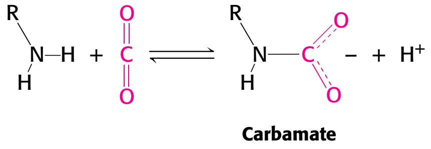


FIGURE 7.22 Carbon dioxide effects. The presence of carbon dioxide decreases the affinity of hemoglobin for oxygen even beyond the effect due to a decrease in pH, resulting in even more efficient oxygen transport from the tissues to the lungs.

In the second mechanism, a direct chemical interaction between carbon dioxide and hemoglobin stimulates oxygen release. The effect of carbon dioxide on oxygen affinity can be seen by comparing oxygen-binding curves in the absence and in the presence of carbon dioxide *at a constant pH* (Figure 7.22). In the presence of carbon dioxide at a partial pressure of 40 torr at pH 7.2, the amount of oxygen released approaches 90% of the maximum carrying capacity. Carbon dioxide stabilizes deoxyhemoglobin by reacting with the terminal amino groups to form *carbamate* groups, which are negatively charged, in contrast with the neutral or positive charges on the free amino groups.



The amino termini lie at the interface between the αβ dimers, and these negatively charged carbamate groups participate in salt-bridge interactions that stabilize the T state, favoring the release of oxygen.

Carbamate formation also provides a mechanism for carbon dioxide transport from tissues to the lungs, but it accounts for only about 14% of the total carbon dioxide transport. Most carbon dioxide released from red blood cells is transported to the lungs in the form of HCO_3^- produced from the hydration of carbon dioxide inside the cell (Figure 7.23). Much of the HCO_3^- that is formed leaves the cell through a specific membrane-transport protein that exchanges HCO_3^- from one side of the membrane for Cl^- from the other side. Thus, the serum concentration of HCO_3^- increases. By this means, a large concentration of carbon dioxide is transported from tissues to the lungs in the form of HCO_3^-. In the lungs, this process is reversed: HCO_3^- is converted back into carbon dioxide and exhaled. Thus, carbon dioxide generated by active tissues contributes to a decrease in red-blood-cell pH and, hence, to oxygen release and is converted into a form that can be transported in the serum and released in the lungs.

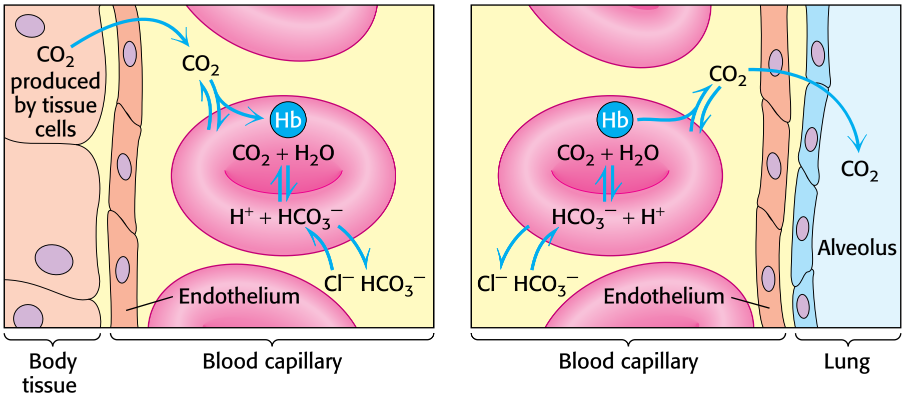


FIGURE 7.23 Transport of CO_2 from tissues to lungs. Most carbon dioxide is transported to the lungs in the form of HCO_3^- produced in red blood cells and then released into the blood plasma. A lesser amount is transported by hemoglobin in the form of an attached carbamate.

7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease

In modern times, particularly after the sequencing of the human genome, it is routine to think of genetically encoded variations in protein sequence as a factor in specific diseases. The notion that diseases might be caused by molecular defects was proposed by Linus Pauling in 1949 (four years before

Watson and Crick's proposal of the DNA double helix) to explain the blood disease *sickle-cell anemia*. The name of the disorder comes from the abnormal sickle shape of red blood cells deprived of oxygen in people suffering from this disease (Figure 7.24). Pauling proposed that sickle-cell anemia might be caused by a specific variation in the amino acid sequence of one hemoglobin chain. Today, we know that this bold hypothesis is correct. In fact, approximately 7% of the world's population are carriers of some disorder of hemoglobin caused by a variation in its amino acid sequence. In concluding this chapter, we will focus on the two most important of these disorders, sickle-cell anemia and thalassemia.

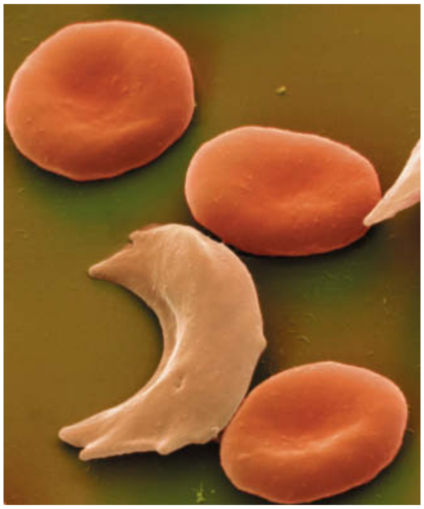


FIGURE 7.24 Sickled red blood cells. A micrograph showing a sickled red blood cell adjacent to normally shaped red blood cells. [Eye of Science/Photo Researchers.]

Sickle-cell anemia results from the aggregation of mutated deoxyhemoglobin molecules

People with sickled red blood cells experience a number of dangerous symptoms. Examination of the contents of these red cells reveals that the hemoglobin molecules have formed large fibrous aggregates (Figure 7.25). These fibers extend across the red blood cells, distorting them so that they clog small capillaries and impair blood flow. In addition, red cells from sickle cell patients are more adherent to the walls of blood vessels than those from normal individuals, prolonging the opportunity for capillary occlusion. The results may be painful swelling of the extremities and a higher risk of stroke or bacterial infection (due to poor circulation). The sickled red cells also do not remain in circulation as long as normal cells do, leading to anemia.

What is the molecular defect associated with sickle-cell anemia? Vernon Ingram demonstrated in 1956 that a single amino acid substitution in the β chain of hemoglobin is responsible—namely, the replacement of a glutamate residue with valine in position 6. The mutated form is referred to as *hemoglobin S* (HbS). In people with sickle-cell anemia, both alleles of the hemoglobin β-chain gene (HbB) are mutated. The HbS substitution substantially decreases the solubility of deoxyhemoglobin, although it does not markedly alter the properties of oxyhemoglobin.

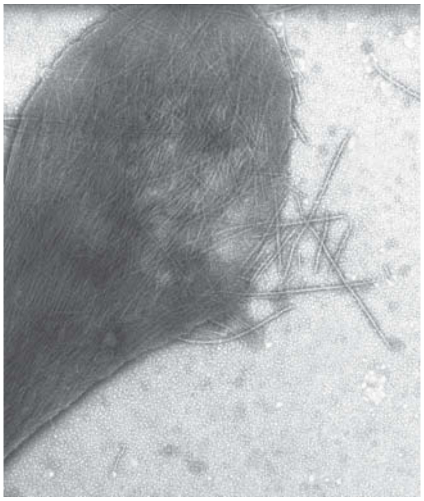


FIGURE 7.25 Sickle-cell hemoglobin fibers. An electron micrograph depicting a ruptured sickled red blood cell with fibers of sickle-cell hemoglobin emerging. [Courtesy of Robert Josephs and Thomas E. Wellems, University of Chicago.]

Examination of the structure of hemoglobin S reveals that the new valine residue lies on the surface of the T-state molecule (Figure 7.26). This new hydrophobic patch interacts with another hydrophobic patch formed by Phe 85 and Leu 88 of the β chain of a neighboring molecule to initiate the aggregation process. More-detailed analysis reveals that a single hemoglobin S fiber is formed from 14 chains of multiple interlinked hemoglobin molecules. Why do these aggregates not form when hemoglobin S is oxygenated? Oxyhemoglobin S is in the R state, and residues Phe 85 and Leu 88 on the β chain are largely buried inside the hemoglobin assembly. In the

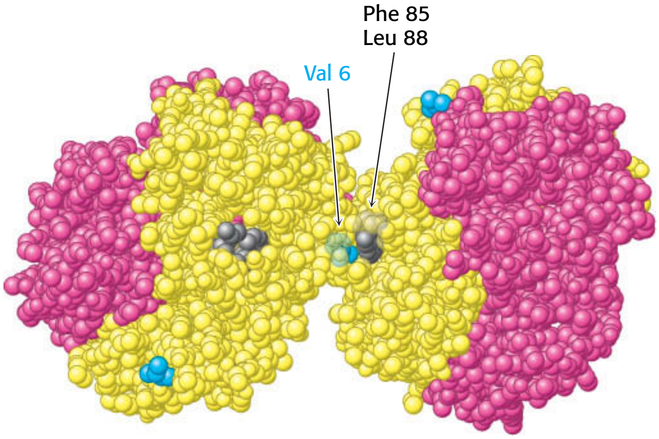


FIGURE 7.26 Deoxygenated hemoglobin S. The interaction between Val 6 (blue) on a β chain of one hemoglobin molecule and a hydrophobic patch formed by Phe 85 and Leu 88 (gray) on a β chain of another deoxygenated hemoglobin molecule leads to hemoglobin aggregation. The exposed Val 6 residues of other β chains participate in other such interactions in hemoglobin S fibers. [Drawn from 2HBS.pdb.]

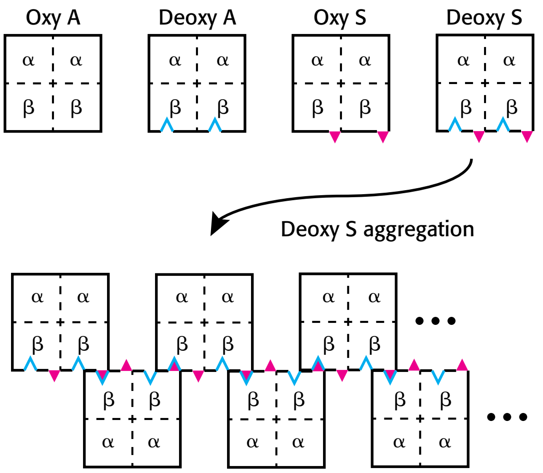
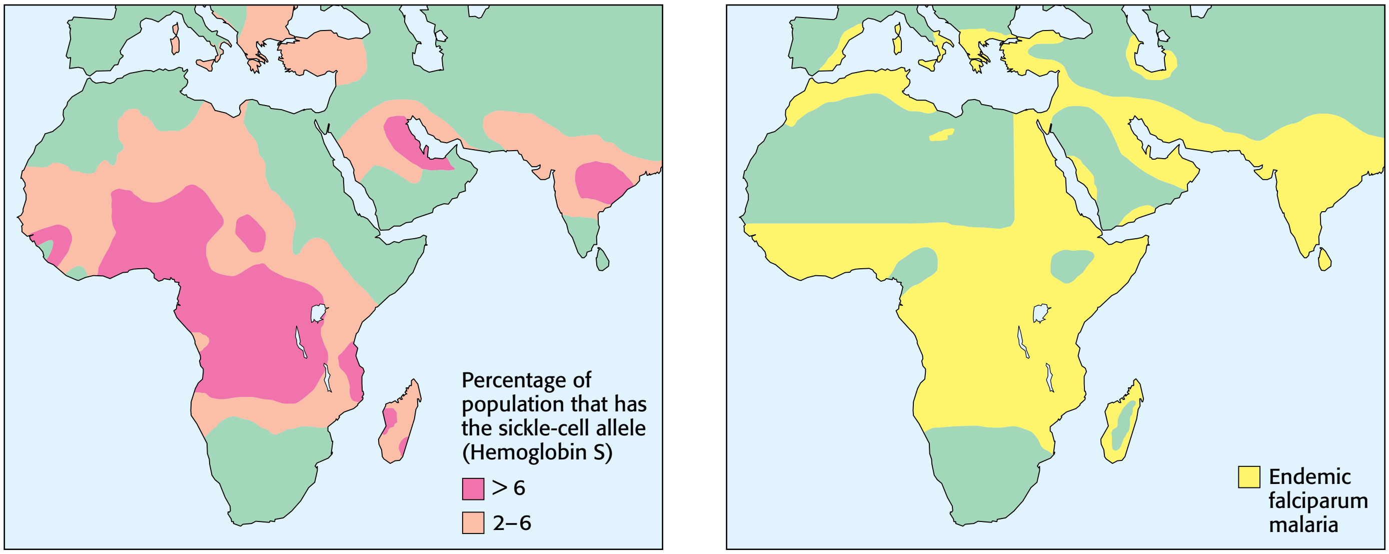


FIGURE 7.27 The formation of HbS aggregates. The mutation to Val 6 in hemoglobin S is represented by the red triangles, while the hydrophobic patch formed by Phe 85 and Leu 88 in deoxyhemoglobin is represented by the blue indentations. When HbS is in its deoxy form, it exhibits the complementary features necessary for aggregation.

absence of a partner with which to interact, the surface Val residue in position 6 is benign (Figure 7.27).

Approximately 1 in 100 West Africans suffer from sickle-cell anemia. Given the often devastating consequences of the disease, why is the HbS mutation so prevalent in Africa and in some other regions? Recall that both copies of the HbB gene are mutated in people with sickle-cell anemia. Individuals with one copy of the HbB gene and one copy of the HbS gene are said to have *sickle-cell trait* because they can pass the HbS gene to their offspring. While sickle-cell trait is considered a benign condition, rare complications have been identified, including an increased risk of exercise-related death in high-performance athletes. However, people with sickle-cell trait exhibit enhanced resistance to *malaria,* a disease carried by a parasite, *Plasmodium falciparum,* that lives within red blood cells at one stage in its life cycle. The dire effect of malaria on health and reproductive likelihood in historically endemic regions has favored people with sickle-cell trait, increasing the prevalence of the HbS allele (Figure 7.28).

FIGURE 7.28 Sickle-cell trait and malaria. A significant correlation is observed between regions with a high frequency of the HbS allele and regions with a high prevalence of malaria.


Percentage of population that has the sickle-cell allele (Hemoglobin S)
> 6
2–6
Endemic falciparum malaria

Thalassemia is caused by an imbalanced production of hemoglobin chains

Sickle-cell anemia is caused by the substitution of a single specific amino acid in one hemoglobin chain. *Thalassemia,* the other prevalent inherited disorder of hemoglobin, is caused by the loss or substantial reduction of a single hemoglobin *chain*. The result is low levels of functional hemoglobin and a decreased production of red blood cells, which may lead to anemia, fatigue, pale skin, and spleen and liver malfunction. Thalassemia is a set of related diseases. In α-thalassemia, the α chain of hemoglobin is not produced in sufficient quantity. Consequently, hemoglobin tetramers form that contain only the β chain. These tetramers, referred to as *hemoglobin H* (HbH), bind oxygen with high affinity and no cooperativity. Thus, oxygen release in the tissues is poor. In β-thalassemia, the β chain of hemoglobin is not produced in sufficient quantity. In the absence of β chains, the α chains form insoluble aggregates that precipitate inside immature red blood cells. The loss of red blood cells results in anemia. The most severe form of β-thalassemia is called *thalassemia major* or *Cooley anemia*.

Both α- and β-thalassemia are associated with many different genetic variations and display a wide range of clinical severity. The most severe forms of α-thalassemia are usually fatal shortly before or just after birth. However, these forms are relatively rare. An examination of the repertoire of hemoglobin genes in the human genome provides one explanation. Normally, humans have not two but four alleles for the α chain, arranged such that the two genes are located adjacent to each other on one end of each chromosome 16. Thus, the complete loss of α-chain expression requires the disruption of four alleles. β-Thalassemia is more common because humans normally have only two alleles for the β chain, one on each copy of chromosome 11.

The accumulation of free alpha-hemoglobin chains is prevented

The presence of four genes expressing the α chain, compared with two for the β chain, suggests that the α chain would be produced in excess (given the overly simple assumption that protein expression from each gene is comparable). If this is correct, why doesn't the excess α chain precipitate? One mechanism for maintaining α chains in solution was revealed by the discovery of an 11-kDa protein in red blood cells called *α-hemoglobin stabilizing protein* (AHSP). This protein forms a soluble complex specifically with newly synthesized α-chain monomers. The crystal structure of a complex between AHSP and α-hemoglobin reveals that AHSP binds to the same face of α-hemoglobin as does β-hemoglobin (Figure 7.29). AHSP

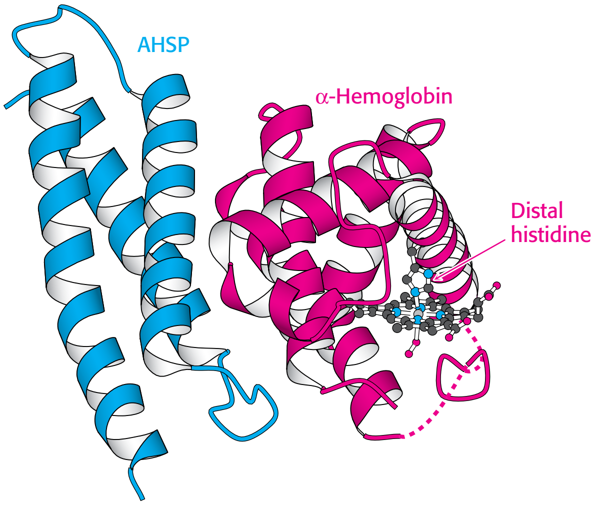


FIGURE 7.29 Stabilizing free α-hemoglobin. The structure of a complex between AHSP and α-hemoglobin is shown. In this complex, the iron atom is bound to oxygen and to the distal histidine. *Notice* that AHSP binds to the same surface of α-hemoglobin as does β-hemoglobin. [Drawn from 1Y01.pdb.]

binds the α chain in both the deoxygenated and oxygenated forms. In the complex with oxygen bound, the distal histidine, rather than the proximal histidine, binds the iron atom.

AHSP serves to bind and ensure the proper folding of α-hemoglobin as it is produced. As β-hemoglobin is expressed, it displaces AHSP because the α-hemoglobin–β-hemoglobin dimer is more stable than the α-hemoglobin–AHSP complex. Thus, AHSP prevents the misfolding, accumulation, and precipitation of free α-hemoglobin. Studies are under way to determine if mutations in the gene encoding AHSP play a role in modulating the severity of β-thalassemia.

Additional globins are encoded in the human genome

In addition to the gene for myoglobin, the two genes for α-hemoglobin, and the one for β-hemoglobin, the human haploid genome contains other globin genes. We have already encountered fetal hemoglobin, which contains the γ chain in place of the β chain. Several other genes encode other hemoglobin subunits that are expressed during development, including the δ chain, the ε chain, and the ζ chain.

Examination of the human genome sequence has revealed two additional globins. Both of these proteins are monomeric proteins, more similar to myoglobin than to hemoglobin. The first, *neuroglobin,* is expressed primarily in the brain and at especially high levels in the retina. Neuroglobin may play a role in protecting neural tissues from hypoxia (insufficient oxygen). The second, *cytoglobin,* is expressed more widely throughout the body. Structural and spectroscopic studies reveal that, in both neuroglobin and cytoglobin, the proximal and the distal histidines are coordinated to the iron atom in the deoxy form. Oxygen binding displaces the distal histidine. Future studies should more completely elucidate the functions of these members of the globin family.

SUMMARY

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme

Myoglobin is a largely α-helical protein that binds the prosthetic group heme. Heme consists of protoporphyrin, an organic component with four linked pyrrole rings, and a central iron ion in the Fe^{2+} state. The iron ion is coordinated to the side chain of a histidine residue in myoglobin, referred to as the proximal histidine. One of the oxygen atoms in O_2 binds to an open coordination site on the iron. Because of partial electron transfer from the iron to the oxygen, the iron ion moves into the plane of the porphyrin on oxygen binding. Hemoglobin consists of four polypeptide chains, two α chains and two β chains. Each of these chains is similar in amino acid sequence to myoglobin and folds into a very similar three-dimensional structure. The hemoglobin tetramer is best described as a pair of αβ dimers.

7.2 Hemoglobin Binds Oxygen Cooperatively

The oxygen-binding curve for myoglobin reveals a simple equilibrium binding process. Myoglobin is half-saturated with oxygen at an oxygen concentration of approximately 2 torr. The oxygen-binding curve for hemoglobin has an "S"-like (sigmoid) shape, indicating that the oxygen binding is cooperative. The binding of oxygen at one site within the hemoglobin tetramer affects the affinities of the other sites for oxygen. Cooperative oxygen binding and release significantly increase the efficiency of oxygen transport. The amount of the potential oxygen-carrying capacity utilized in transporting oxygen from the lungs (with a partial pressure of oxygen of 100 torr) to tissues (with a partial pressure of oxygen of 20 torr) is 66% compared with 7% if myoglobin had been used as the oxygen carrier.

The quaternary structure of hemoglobin changes on oxygen binding. The structure of deoxyhemoglobin is referred to as the T state. The structure of oxyhemoglobin is referred to as the R state. The two αβ dimers rotate by approximately 15 degrees with respect to one another in the transition from the T to the R state. Cooperative binding can be potentially explained by concerted and sequential models. In the concerted model, each hemoglobin adopts either the T state or the R state; the equilibrium between these two states is determined by the number of occupied oxygen-binding sites. Sequential models allow intermediate structures. Structural changes at the iron sites in response to oxygen binding are transmitted to the interface between αβ dimers, influencing the T-to-R equilibrium.

Red blood cells contain 2,3-bisphosphoglycerate in concentrations approximately equal to that for hemoglobin. 2,3-BPG binds tightly to the T state but not to the R state, stabilizing the T state and lowering the oxygen affinity of hemoglobin. Fetal hemoglobin binds oxygen more tightly than does adult hemoglobin owing to weaker 2,3-BPG binding. This difference allows oxygen transfer from maternal to fetal blood.

7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen: The Bohr Effect

The oxygen-binding properties of hemoglobin are markedly affected by pH and by the presence of carbon dioxide, a phenomenon known as the Bohr effect. Increasing the concentration of hydrogen ions—that is, decreasing pH—decreases the oxygen affinity of hemoglobin, owing to the protonation of the amino termini and certain histidine residues. The protonated residues help stabilize the T state. Increasing concentrations of carbon dioxide decrease the oxygen affinity of hemoglobin by two mechanisms. First, carbon dioxide is converted into carbonic acid, which lowers the oxygen affinity of hemoglobin by decreasing the pH inside the red blood cell. Second, carbon dioxide adds to the amino termini of hemoglobin to form carbamates. These negatively charged groups stabilize deoxyhemoglobin through ionic interactions. Because hydrogen ions and carbon dioxide are produced in rapidly metabolizing tissues, the Bohr effect helps deliver oxygen to sites where it is most needed.

7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease

Sickle-cell disease is caused by a mutation in the β chain of hemoglobin that substitutes a valine residue for a glutamate residue. As a result, a hydrophobic patch forms on the surface of deoxy (T-state) hemoglobin that leads to the formation of fibrous polymers. These fibers distort red blood cells into sickle shapes. Sickle-cell disease was the first disease to be associated with a change in the amino acid sequence of a protein. Thalassemias are diseases caused by the reduced production of either the α or the β chain, yielding hemoglobin tetramers that contain only one type of hemoglobin chain. Such hemoglobin molecules are characterized by poor oxygen release and low solubility, leading to the destruction of red blood cells in the course of their development.

Red-blood-cell precursors normally produce a slight excess of hemoglobin α chains compared with β chains. To prevent the aggregation of the excess α chains, they produce α-hemoglobin stabilizing protein, which binds specifically to newly synthesized α-chain monomers to form a soluble complex.

APPENDIX: Binding Models Can Be Formulated in Quantitative Terms: The Hill Plot and the Concerted Model

The Hill Plot

A useful way of quantitatively describing cooperative binding processes such as that for hemoglobin was developed by Archibald Hill in 1913. Consider the *hypothetical* equilibrium for a protein X binding a ligand S:

$$\mathrm{X} + n\mathrm{S} \rightleftharpoons \mathrm{X(S)}_n \qquad (1)$$

where n is a variable that can take on both integral and fractional values. The parameter n is a measure of the degree of cooperativity in ligand binding. For X = hemoglobin and S = O_2, the maximum value of n is 4. The value of $n = 4$ would apply if oxygen binding by hemoglobin were completely cooperative. If oxygen binding were completely noncooperative, then n would be 1.

Analysis of the equilibrium in equation 1 yields the following expression for the fractional saturation, Y:

$$Y = \frac{[\mathrm{S}]^n}{[\mathrm{S}]^n + [\mathrm{S}_{50}]^n}$$

where $[\mathrm{S}_{50}]$ is the concentration at which X is half-saturated. For hemoglobin, this expression becomes

$$Y = \frac{p\mathrm{O}_2{}^n}{p\mathrm{O}_2{}^n + P_{50}{}^n}$$

where P_{50} is the partial pressure of oxygen at which hemoglobin is half-saturated. This expression can be rearranged to:

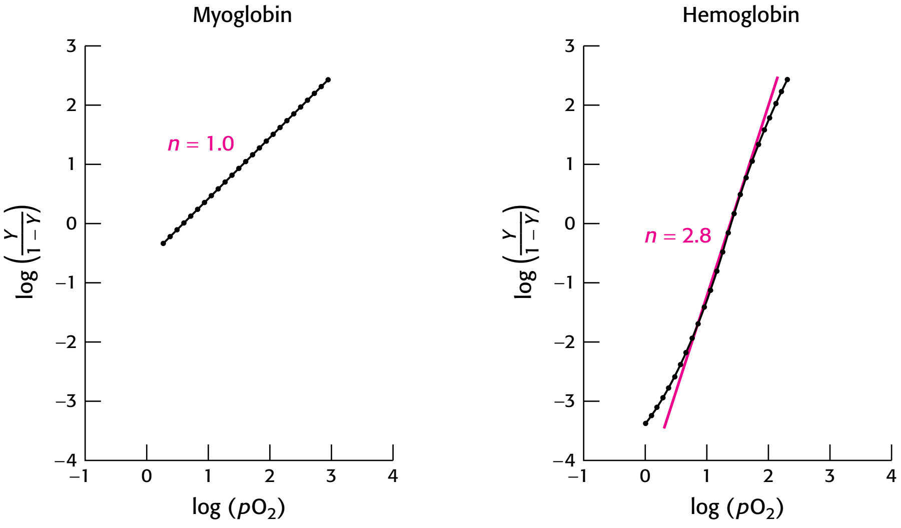
$$\frac{Y}{1 - Y} = \frac{p\mathrm{O}_2{}^n}{P_{50}{}^n}$$

and so

$$\log\left(\frac{Y}{1 - Y}\right) = \log\left(\frac{p\mathrm{O}_2{}^n}{P_{50}{}^n}\right) = n\log(p\mathrm{O}_2) - n\log(P_{50})$$

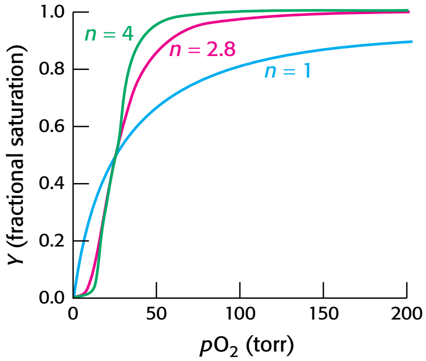
This equation predicts that a plot of $\log(Y/1 - Y)$ versus $\log(P_{50})$, called a *Hill plot*, should be linear with a slope of n.

Hill plots for myoglobin and hemoglobin are shown in Figure 7.30. For myoglobin, the Hill plot is linear with a slope of 1. For hemoglobin, the Hill plot is not completely linear, because the equilibrium on which the Hill plot is based is not entirely correct. However, the plot is approximately linear in the center with a slope of 2.8. The slope, often referred to as the *Hill coefficient,* is a measure of the cooperativity of oxygen binding. The utility of the Hill plot is that it provides a simply derived quantitative assessment of the degree of cooperativity in binding. With the use of the Hill equation and the derived Hill coefficient, a binding



APPENDIX FIGURE 7.30 Hill plots for myoglobin and hemoglobin.

curve that closely resembles that for hemoglobin is produced (Figure 7.31).



APPENDIX FIGURE 7.31 Oxygen-binding curves for several Hill coefficients. The curve labeled $n = 2.8$ closely resembles the curve for hemoglobin.

The Concerted Model

The concerted model can be formulated in quantitative terms. Only four parameters are required: (1) the number of binding sites (assumed to be equivalent) in the protein, (2) the ratio of the concentrations of the T and R states in the absence of bound ligands, (3) the affinity of sites in proteins in the R state for ligand binding, and (4) a measure of how much more tightly subunits in proteins in the R state bind ligands compared with subunits in the T state. The number of binding sites, n, is usually known from other information. For hemoglobin, $n = 4$. The ratio of the concentrations of the T and R states with no ligands bound, L, is the allosteric constant:

$$L = [T_0]/[R_0]$$

where the subscript refers to the number of ligands bound (in this case, zero). The affinity of subunits in the R state is defined by the dissociation constant for a ligand binding to a single site in the R state, K_R. Similarly, the dissociation constant for a ligand binding to a single site in the T state is K_T. We can define the ratio of these two dissociation constant as

$$c = K_R/K_T$$

This is the measure of how much more tightly a subunit for a protein in the R state binds a ligand compared with a subunit for a protein in the T state. Note that $c < 1$ because K_R and K_T are dissociation constants and tight binding corresponds to a small dissociation constant.

What is the ratio of the concentration of T-state proteins with one ligand bound to the concentration of R-state proteins with one ligand bound? The dissociation constant for a single site in the R state is K_R. For a protein with n sites, there are n possible sites for the first ligand to bind. This statistical factor favors ligand binding compared with a single-site protein. Thus, $[R_1] = n[R_0][S]/K_R$. Similarly, $[T_1] = n[T_0][S]/K_T$. Thus,

$$[T_1]/[R_1] = \frac{n[T_0][S]/K_T}{n[R_0][S]/K_R} = \frac{[T_0]}{[R_0](K_R/K_T)} = cL$$

Similar analysis reveals that, for states with i ligands bound, $[T_i]/[R_i] = c^iL$. In other words, the ratio of the concentrations of the T state to the R state is reduced by a factor of c for each ligand that binds.

Let us define a convenient scale for the concentration of S:

$$\alpha = [S]/K_R$$

This definition is useful because it is the ratio of the concentration of S to the dissociation constant that determines the extent of binding. Using this definition, we see that

$$[R_1] = \frac{n[R_0][S]}{K_R} = n[R_0]\alpha$$

Similarly,

$$[T_1] = \frac{n[T_0][S]}{K_T} = ncL[R_0]\alpha$$

What is the concentration of R-state molecules with two ligands bound? Again, we must consider the statistical factor—that is, the number of ways in which a second ligand can bind to a molecule with one site occupied. The number of ways is $n - 1$. However, because which ligand is the "first" and which is the "second" does not matter, we must divide by a factor of 2. Thus,

$$\begin{aligned}[R_2] &= \frac{\left(\frac{n-1}{2}\right)[R_1][S]}{K_R} \\ &= \left(\frac{n-1}{2}\right)[R_1]\alpha \\ &= \left(\frac{n-1}{2}\right)(n[R_0]\alpha)\alpha \\ &= n\left(\frac{n-1}{2}\right)[K_0]\alpha^2\end{aligned}$$

We can derive similar equations for the case with i ligands bound and for T states.

We can now calculate the fractional saturation, Y. This is the total concentration of sites with ligands bound divided by the total concentration of potential binding sites. Thus,

$$Y = \frac{([R_1] + [T_1] + 2([R_2]) + [T_2]) + \cdots + n([R_n] + [T_n])}{n([R_0] + [T_0] + [R_1] + [T_1] + \cdots + [R_n] + [T_n])}$$

Substituting into this equation, we find

$$Y = \frac{\begin{array}{r} n[\mathrm{R}_0]\alpha + nc[\mathrm{T}_0]\alpha + 2(n(n-1)/2)[\mathrm{R}_0]\alpha^2 \\ + 2(n(n-1)/2)c^2[\mathrm{T}_0]\alpha^2 + \cdots \\ + n[\mathrm{R}_0]\alpha^n + nc^n[\mathrm{T}^0])\alpha^n \end{array}}{\begin{array}{r} n([\mathrm{R}_0] + [\mathrm{T}_0] + n[\mathrm{R}_0]\alpha + nc[\mathrm{T}_0]\alpha + \cdots \\ + [\mathrm{R}_0]\alpha^n + c^n[\mathrm{T}_0]\alpha^n) \end{array}}$$

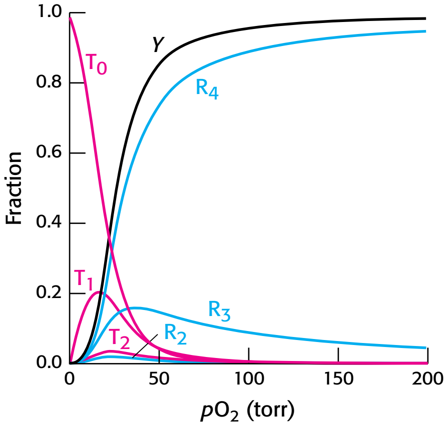
Substituting $[\mathrm{T}_0] = L[\mathrm{R}_0]$ and summing these series yields

$$Y = \frac{\alpha(1+\alpha)^{n-1} + Lc\alpha(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n}$$

We can now use this equation to fit the observed data for hemoglobin by varying the parameters L, c, and K_R (with $n = 4$). An excellent fit is obtained with $L = 9000$, $c = 0.014$, and $K_R = 2.5$ torr (Figure 7.32).

In addition to the fractional saturation, the concentrations of the species T_0, T_1, T_2, R_2, R_3, and R_4 are shown. The concentrations of all other species are very low. The addition of concentrations is a major difference between the analysis using the Hill equation and this analysis of the concerted model. The Hill equation gives only the fractional saturation, whereas the analysis of the concerted model yields concentrations for all species. In the present case, this analysis yields the expected ratio of T-state proteins to R-state proteins at each stage of binding. This ratio changes from 9000 to 126 to 1.76 to 0.025 to 0.00035 with zero, one, two, three, and four oxygen molecules bound. This ratio provides a quantitative measure of the switching of the population of hemoglobin molecules from the T state to the R state.

The sequential model can also be formulated in quantitative terms. However, the formulation entails many more parameters, and many different sets of parameters often yield similar fits to the experimental data.



APPENDIX FIGURE 7.32 Modeling oxygen binding with the concerted model. The fractional saturation (γ) as a function pO_2: $L = 9000$, $c = 0.014$, and $K_R = 2.5$ torr. The fraction of molecules in the T state with zero, one, and two oxygen molecules bound (T_0, T_1, and T_2) and the fraction of molecules in the R state with two, three, and four oxygen molecules bound (R_2, R_3, and R_4) are shown. The fractions of molecules in other forms are too low to be shown.

KEY TERMS

heme (p. 192)
protoporphyrin (p. 192)
proximal histidine (p. 193)
functional magnetic resonance imaging (fMRI) (p. 193)
superoxide anion (p. 194)
metmyoglobin (p. 194)
distal histidine (p. 194)
α chain (p. 195)
β chain (p. 195)
globin fold (p. 195)
αβ dimer (p. 195)
oxygen-binding curve (p. 195)
fractional saturation (p. 195)
partial pressure (p. 195)
sigmoid (p. 196)
cooperative binding (p. 196)
T state (p. 198)
R state (p. 198)
concerted model (MWC model) (p. 198)
sequential model (p. 199)
2,3-bisphosphoglycerate (p. 200)
fetal hemoglobin (p. 201)
carbon monoxide (p. 201)
carboxyhemoglobin (p. 201)
Bohr effect (p. 202)
carbonic anhydrase (p. 203)
carbamate (p. 204)
sickle-cell anemia (p. 205)
hemoglobin S (p. 205)
sickle-cell trait (p. 206)
malaria (p. 206)
thalassemia (p. 207)
hemoglobin H (p. 207)
thalassemia major (Cooley anemia) (p. 207)
α-hemoglobin stabilizing protein (AHSP) (p. 207)
neuroglobin (p. 208)
cytoglobin (p. 208)
Hill plot (p. 210)
Hill coefficient (p. 210)

PROBLEMS

1. *Screening the biosphere.* The first protein structure to have its structure determined was myoglobin from sperm whale. Propose an explanation for the observation that sperm whale muscle is a rich source of this protein.

2. *Hemoglobin content.* The average volume of a red blood cell is 87 μm^3. The mean concentration of hemoglobin in red cells is 0.34 g ml^{-1}.

(a) What is the weight of the hemoglobin contained in an average red cell?

(b) How many hemoglobin molecules are there in an average red cell? Assume that the molecular weight of the human hemoglobin tetramer is 65 kDa.

(c) Could the hemoglobin concentration in red cells be much higher than the observed value? (Hint: Suppose that a red cell contained a crystalline array of hemoglobin molecules in a cubic lattice with 65 Å sides.)

3. *Iron content.* How much iron is there in the hemoglobin of a 70-kg adult? Assume that the blood volume is 70 ml kg^{-1} of body weight and that the hemoglobin content of blood is 0.16 g ml^{-1}.

4. *Oxygenating myoglobin.* The myoglobin content of some human muscles is about 8 g kg^{-1}. In sperm whale, the myoglobin content of muscle is about 80 g kg^{-1}.

(a) How much O_2 is bound to myoglobin in human muscle and in sperm whale muscle? Assume that the myoglobin is saturated with O_2, and that the molecular weights of human and sperm whale myoglobin are the same.

(b) The amount of oxygen dissolved in tissue water (in equilibrium with venous blood) at 37°C is about 3.5×10^{-5} M. What is the ratio of oxygen bound to myoglobin to that directly dissolved in the water of sperm whale muscle?

5. *Tuning proton affinity.* The pK_a of an acid depends partly on its environment. Predict the effect of each of the following environmental changes on the pK_a of a glutamic acid side chain.

(a) A lysine side chain is brought into proximity.

(b) The terminal carboxyl group of the protein is brought into proximity.

(c) The glutamic acid side chain is shifted from the outside of the protein to a nonpolar site inside.

6. *Saving grace.* Hemoglobin A inhibits the formation of the long fibers of hemoglobin S and the subsequent sickling of the red cell on deoxygenation. Why does hemoglobin A have this effect?

7. *Carrying a load.* Suppose that you are climbing a high mountain and the oxygen partial pressure in the air is reduced to 75 torr. Estimate the percentage of the oxygen-carrying capacity that will be utilized, assuming that the pH of both tissues and lungs is 7.4 and that the oxygen concentration in the tissues is 20 torr.

8. *Bohr for me, not for thee.* Does myoglobin exhibit a Bohr effect? Why or why not?

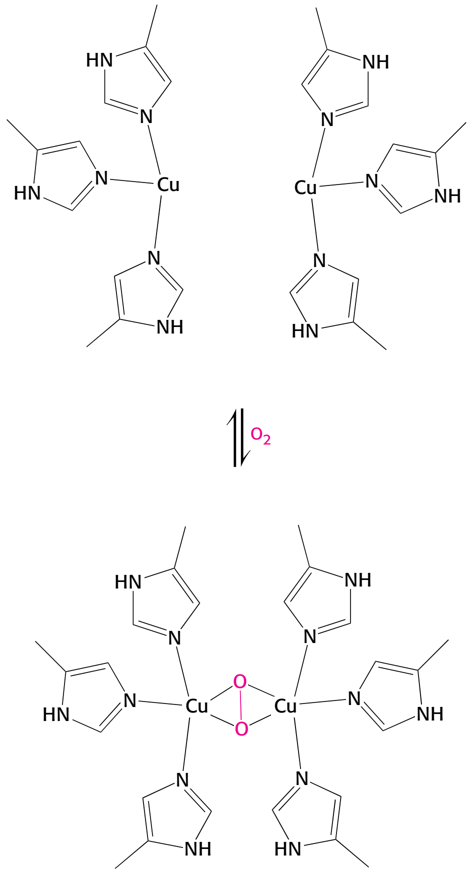
9. *High-altitude adaptation.* After spending a day or more at high altitude (with an oxygen partial pressure of 75 torr), the concentration of 2,3-bisphosphoglycerate (2,3-BPG) in red blood cells increases. What effect would an increased concentration of 2,3-BPG have on the oxygen-binding curve for hemoglobin? Why would this adaptation be beneficial for functioning well at high altitude?

10. *Blood doping.* Endurance athletes sometimes try an illegal method of blood doping called autologous transfusion. Some blood from the athlete is removed well before competition, and then transfused back into the athlete just before competition.

(a) Why might blood transfusion benefit the athlete?

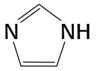
(b) With time, stored red blood cells become depleted in 2,3-BPG. What might be the consequences of using such blood for a blood transfusion?

11. *I'll take the lobster.* Arthropods such as lobsters have oxygen carriers quite different from hemoglobin. The oxygen-binding sites do not contain heme but, instead, are based on two copper(I) ions. The structural changes that accompany oxygen binding are shown below. How might these changes be used to facilitate cooperative oxygen binding?



12. *A disconnect.* With the use of site-directed mutagenesis, hemoglobin has been prepared in which the proximal histidine residues in both the α and the β subunits have been replaced by glycine. The imidazole ring from the histidine residue can be replaced by adding free imidazole in

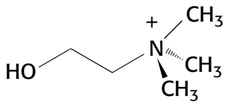
solution. Would you expect this modified hemoglobin to show cooperativity in oxygen binding? Why or why not?



Imidazole

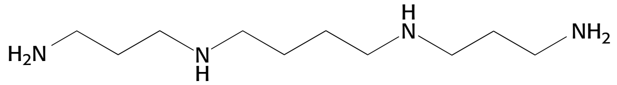
13. *Successful substitution*. Blood cells from some birds do not contain 2,3-bisphosphoglycerate but, instead, contain one of the compounds in parts *a* through *d*, which plays an analogous functional role. Which compound do you think is most likely to play this role? Explain briefly.

(a)



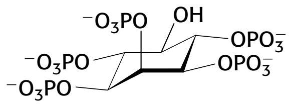
Choline

(b)



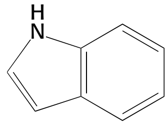
Spermine

(c)



Inositol pentaphosphate

(d)



Indole

14. *Theoretical curves*. (a) Using the Hill equation, plot an oxygen-binding curve for a hypothetical two-subunit hemoglobin with $n = 1.8$ and $P_{50} = 10$ torr. (b) Repeat, using the concerted model with $n = 2$, $L = 1000$, $c = 0.01$, and $K_R = 1$ torr.

15. *Parasitic effect*. When *P. falciparum* lives inside red blood cells, the metabolism of the parasite tends to release acid. What effect is the presence of acid likely to have on the oxygen-carrying capacity of the red blood cells? On the likelihood that these cells sickle?

Data Interpretation Problems

16. *Primitive oxygen binding*. Lampreys are primitive organisms whose ancestors diverged from the ancestors of fish and mammals approximately 400 million years ago. Lamprey blood contains a hemoglobin related to mammalian hemoglobin. However, lamprey hemoglobin is monomeric in the oxygenated state. Oxygen-binding data for lamprey hemoglobin are as follows:

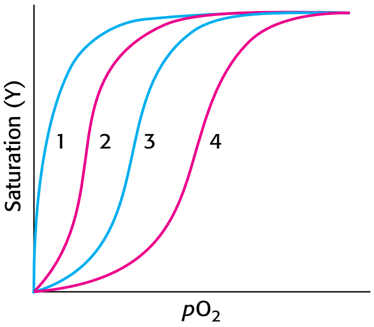
pO_2	Y	pO_2	Y	pO_2	Y
0.1	.0060	2.0	.112	50.0	.889
0.2	.0124	3.0	.170	60.0	.905
0.3	.0190	4.0	.227	70.0	.917
0.4	.0245	5.0	.283	80.0	.927
0.5	.0307	7.5	.420	90.0	.935
0.6	.0380	10.0	.500	100	.941
0.7	.0430	15.0	.640	150	.960
0.8	.0481	20.0	.721	200	.970
0.9	.0530	30.0	.812		
1.0	.0591	40.0	.865		

(a) Plot these data to produce an oxygen-binding curve. At what oxygen partial pressure is this hemoglobin half-saturated? On the basis of the appearance of this curve, does oxygen binding seem to be cooperative?

(b) Construct a Hill plot using these data. Does the Hill plot show any evidence for cooperativity? What is the Hill coefficient?

(c) Further studies revealed that lamprey hemoglobin forms oligomers, primarily dimers, in the deoxygenated state. Propose a model to explain any observed cooperativity in oxygen binding by lamprey hemoglobin.

17. *Leaning to the left or to the right*. The illustration below shows several oxygen-dissociation curves. Assume that curve 3 corresponds to hemoglobin with physiological concentrations of CO_2 and 2,3-BPG at pH 7. Which curves represent each of the following perturbations?



(a) Decrease in CO_2

(b) Increase in 2,3-BPG

(c) Increase in pH

(d) Loss of quaternary structure

Chapter Integration Problems

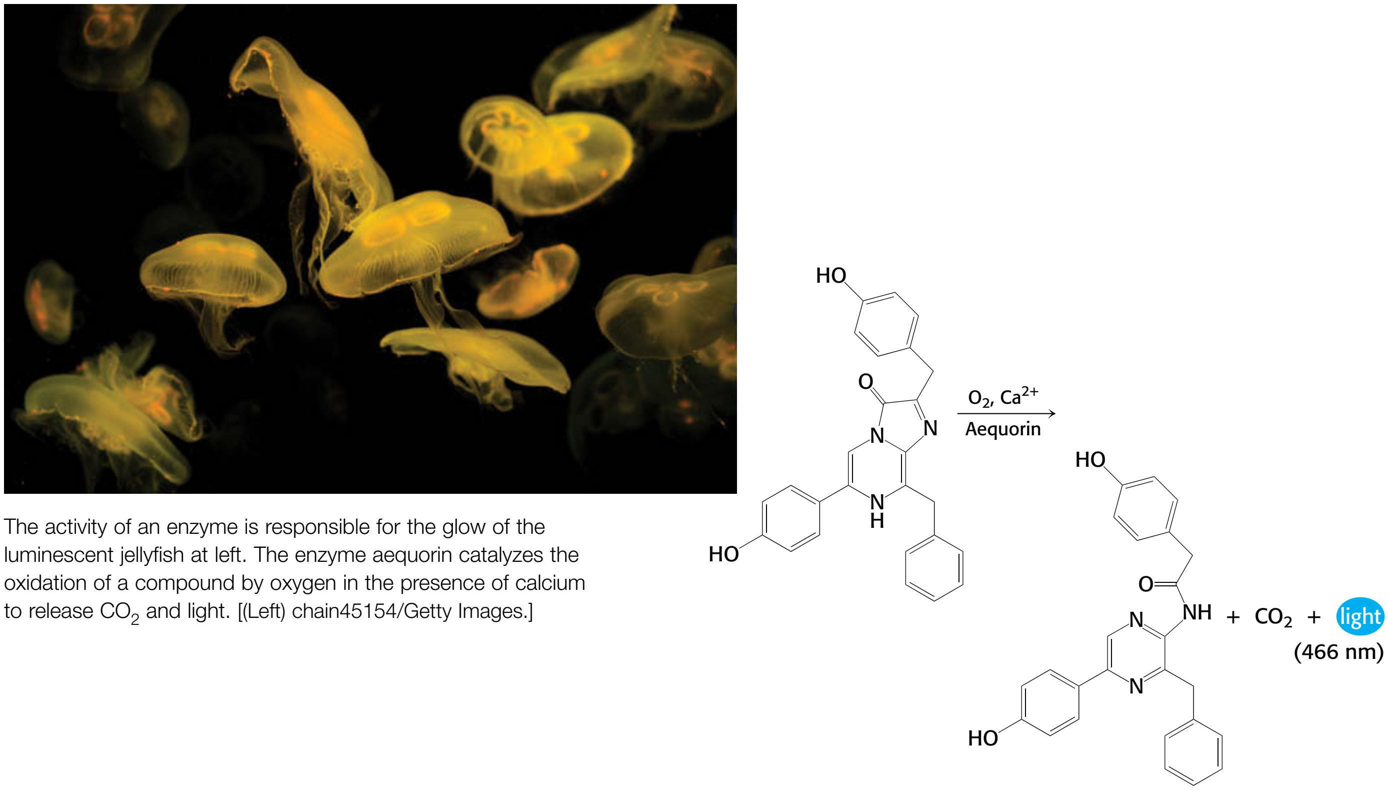
18. *Location is everything*. 2,3-Bisphosphoglycerate lies in a central cavity within the hemoglobin tetramer, stabilizing the T state. What would be the effect of mutations that placed the BPG-binding site on the surface of hemoglobin?

19. *A therapeutic option*. Hydroxyurea has been shown to increase the expression of fetal hemoglobin in adult red blood cells, by a mechanism that remains unclear. Explain why hydroxyurea can be a useful therapy for patients with sickle-cell anemia.

CHAPTER

8

Enzymes: Basic Concepts and Kinetics



The activity of an enzyme is responsible for the glow of the luminescent jellyfish at left. The enzyme aequorin catalyzes the oxidation of a compound by oxygen in the presence of calcium to release CO_2 and light. [(Left) chain45154/Getty Images.]

OUTLINE

8.1 Enzymes Are Powerful and Highly Specific Catalysts

8.2 Gibbs Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes

8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

8.4 The Michaelis–Menten Model Accounts for the Kinetic Properties of Many Enzymes

8.5 Enzymes Can Be Inhibited by Specific Molecules

8.6 Enzymes Can Be Studied One Molecule at a Time

Enzymes, the catalysts of biological systems, are remarkable molecular devices that determine the patterns of chemical transformations. They also mediate the transformation of one form of energy into another. About a quarter of the genes in the human genome encode enzymes, a testament to their importance to life. The most striking characteristics of enzymes are their *catalytic power* and *specificity.* Catalysis takes place at a particular site on the enzyme called the *active site.*

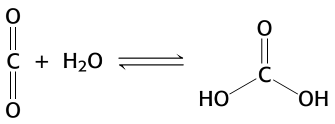
Nearly all known enzymes are proteins. However, proteins do not have an absolute monopoly on catalysis; the discovery of catalytically active RNA molecules provides compelling evidence that RNA was a biocatalyst early in evolution. Proteins as a class of macromolecules are highly effective catalysts for an enormous diversity of chemical reactions because of their capacity *to specifically bind a very wide range of molecules.* By utilizing the full repertoire of intermolecular forces, enzymes bring substrates together in an optimal orientation, the prelude to making and breaking chemical bonds. They catalyze reactions *by stabilizing transition states,* the highest-energy species in reaction pathways. By selectively stabilizing a transition state, an enzyme determines which one of several potential chemical reactions actually takes place.

TABLE 8.1 Rate enhancement by selected enzymes

Enzyme	Nonenzymatic half-life	Uncatalyzed rate (k_{un} s^{-1})	Catalyzed rate (k_{cat} s^{-1})	Rate enhancement (k_{cat} s^{-1}/k_{un} s^{-1})
OMP decarboxylase	78,000,000 years	2.8×10^{-16}	39	1.4×10^{17}
Staphylococcal nuclease	130,000 years	1.7×10^{-13}	95	5.6×10^{14}
AMP nucleosidase	69,000 years	1.0×10^{-11}	60	6.0×10^{12}
Carboxypeptidase A	7.3 years	3.0×10^{-9}	578	1.9×10^{11}
Ketosteroid isomerase	7 weeks	1.7×10^{-7}	66,000	3.9×10^{11}
Triose phosphate isomerase	1.9 days	4.3×10^{-6}	4,300	1.0×10^{9}
Chorismate mutase	7.4 hours	2.6×10^{-5}	50	1.9×10^{6}
Carbonic anhydrase	5 seconds	1.3×10^{-1}	1×10^{6}	7.7×10^{6}

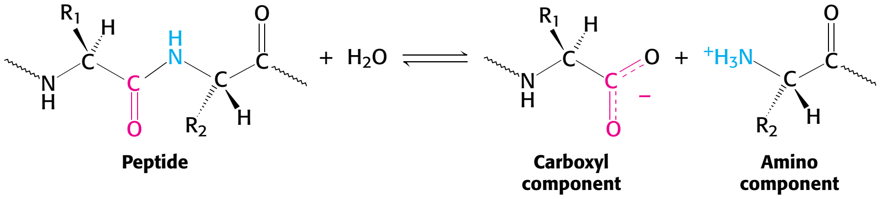
Abbreviations: OMP, orotidine monophosphate; AMP, adenosine monophosphate.
Source: After A. Radzicka and R. Wolfenden. *Science* 267:90–93, 1995.

8.1 Enzymes Are Powerful and Highly Specific Catalysts

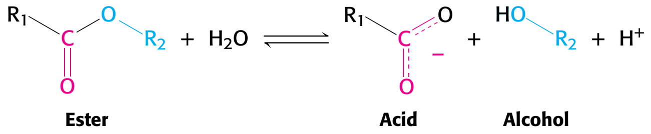


Enzymes accelerate reactions by factors of as much as a million or more (Table 8.1). Indeed, most reactions in biological systems do not take place at perceptible rates in the absence of enzymes. Even a reaction as simple as the hydration of carbon dioxide is catalyzed by an enzyme—namely, carbonic anhydrase. The transfer of CO_2 from the tissues to the blood and then to the air in the alveolae of the lungs would be less complete in the absence of this enzyme (p. 203). In fact, carbonic anhydrase is one of the fastest enzymes known. Each enzyme molecule can hydrate 10^6 molecules of CO_2 *per second*. This catalyzed reaction is 10^7 times as fast as the uncatalyzed one. We will consider the mechanism of carbonic anhydrase catalysis in Chapter 9.

Enzymes are highly specific both in the reactions that they catalyze and in their choice of reactants, which are called *substrates*. An enzyme usually catalyzes a single chemical reaction or a set of closely related reactions. Let us consider *proteolytic enzymes* as an example. The biochemical function of these enzymes is to catalyze *proteolysis,* the hydrolysis of a peptide bond.



Most proteolytic enzymes also catalyze a different but related reaction in vitro—namely, the hydrolysis of an ester bond. Such reactions are more easily monitored than is proteolysis and are useful in experimental investigations of these enzymes.



Proteolytic enzymes differ markedly in their degree of substrate specificity. Papain, which is found in papaya plants, is quite undiscriminating: it will cleave any peptide bond with little regard to the identity of the adjacent side chains. This lack of specificity accounts for its use in meat-tenderizing sauces. The digestive enzyme trypsin, on the other hand, is quite specific and catalyzes the splitting of peptide bonds only on the carboxyl side of lysine and arginine residues (Figure 8.1A). Thrombin, an enzyme that participates in blood clotting (Section 10.4), is even more specific than trypsin. It catalyzes the hydrolysis of Arg–Gly bonds in particular peptide sequences only (Figure 8.1B).

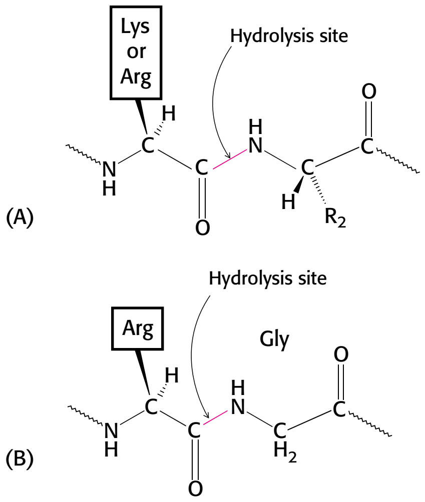


FIGURE 8.1 Enzyme specificity. (A) Trypsin cleaves on the carboxyl side of arginine and lysine residues, whereas (B) thrombin cleaves Arg–Gly bonds in particular sequences only.

DNA polymerase I, a template-directed enzyme (Section 28.3), is another highly specific catalyst. DNA polymerase adds nucleotides to the strand being synthesized in a sequence determined by the sequence of nucleotides in another DNA strand that serves as a template. DNA polymerase I is remarkably precise in carrying out the instructions given by the template. It inserts the wrong nucleotide into a new DNA strand less than one in a thousand times. *The specificity of an enzyme is due to the precise interaction of the substrate with the enzyme. This precision is a result of the intricate three-dimensional structure of the enzyme protein.*

Many enzymes require cofactors for activity

The catalytic activity of many enzymes depends on the presence of small molecules termed *cofactors,* although the precise role varies with the cofactor and the enzyme. Generally, these cofactors are able to execute chemical reactions that cannot be performed by the standard set of twenty amino acids. An enzyme without its cofactor is referred to as an *apoenzyme;* the complete, catalytically active enzyme is called a *holoenzyme*.

$$\text{Apoenzyme} + \text{cofactor} = \text{holoenzyme}$$

Cofactors can be subdivided into two groups: (1) metals and (2) small organic molecules called *coenzymes* (Table 8.2). Often derived from vitamins, coenzymes can be either tightly or loosely bound to the enzyme. Tightly bound coenzymes are called *prosthetic groups*. Loosely associated coenzymes are more like cosubstrates because, like substrates and products, they bind to the enzyme and are released from it. The use of the same coenzyme by a variety of enzymes sets coenzymes apart from normal substrates, however, as does their source in vitamins (Section 15.4). Enzymes that use the same coenzyme usually perform catalysis by similar mechanisms. In Chapter 9, we will examine the importance of metals to enzyme activity and, throughout the book, we will see how coenzymes and their enzyme partners operate in their biochemical context.

TABLE 8.2 Enzyme cofactors

Cofactor	Enzyme
Coenzyme	
Thiamine pyrophosphate	Pyruvate dehydrogenase
Flavin adenine nucleotide	Monoamine oxidase
Nicotinamide adenine dinucleotide	Lactate dehydrogenase
Pyridoxal phosphate	Glycogen phosphorylase
Coenzyme A (CoA)	Acetyl CoA carboxylase
Biotin	Pyruvate carboxylase
5′-Deoxyadenosyl cobalamin	Methylmalonyl mutase
Tetrahydrofolate	Thymidylate synthase
Metal	
Zn^{2+}	Carbonic anhydrase
Zn^{2+}	Carboxypeptidase
Mg^{2+}	*Eco*RV
Mg^{2+}	Hexokinase
Ni^{2+}	Urease
Mo	Nitrogenase
Se	Glutathione peroxidase
Mn	Superoxide dismutase
K^{+}	Acetyl CoA thiolase

Enzymes can transform energy from one form into another

A key activity in all living systems is the conversion of one form of energy into another. For example, in photosynthesis, light energy is converted into chemical-bond energy. In cellular respiration, which takes place in mitochondria, the free energy contained in small molecules derived from food is converted first into the free energy of an ion gradient and then into a different currency—the free energy of adenosine triphosphate. Given their centrality to life, it should come as no surprise that enzymes play vital roles in energy transformation. As we will see, enzymes perform fundamental roles in photosynthesis and cellular respiration. Other enzymes can then use the chemical-bond energy of ATP in diverse ways. For instance, the enzyme myosin converts the energy of ATP into the mechanical energy of contracting muscles (Section 9.4 and Chapter 35). Pumps in the membranes of cells and organelles, which can be thought of as enzymes that move substrates rather than chemically alter them, use the energy of ATP to transport molecules and ions across the membrane (Chapter 13). The chemical and electrical gradients resulting from the unequal distribution of these molecules and ions are themselves forms of energy that can be used for a variety of purposes, such as sending nerve impulses.

The molecular mechanisms of these energy-transducing enzymes are being unraveled. We will see in subsequent chapters how unidirectional cycles of discrete steps—binding, chemical transformation, and release—lead to the conversion of one form of energy into another.

8.2 Gibbs Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes

Enzymes speed up the rate of chemical reactions, but the properties of the reaction—whether it can take place at all and the degree to which the enzyme accelerates the reaction—depend on energy differences between reactants and products. *Gibbs free energy (G),* which was touched on in Chapter 1, is a thermodynamic property that is a measure of useful energy, or the energy that is capable of doing work. To understand how enzymes operate, we need to consider only two thermodynamic properties of the reaction: (1) the free-energy difference (ΔG) between the products and reactants and (2) the energy required to initiate the conversion of reactants into products. The former determines whether the reaction will take place spontaneously, whereas the latter determines the rate of the reaction. Enzymes affect only the latter. Let us review some of the principles of thermodynamics as they apply to enzymes.

The free-energy change provides information about the spontaneity but not the rate of a reaction

As discussed in Chapter 1, the free-energy change of a reaction (ΔG) tells us if the reaction can take place spontaneously:

1. *A reaction can take place spontaneously only if ΔG is negative.* Such reactions are said to be *exergonic*.

2. *A system is at equilibrium and no* net *change can take place if ΔG is zero.*

3. *A reaction cannot take place spontaneously if ΔG is positive.* An input of free energy is required to drive such a reaction. These reactions are termed *endergonic*.

4. *The ΔG of a reaction depends only on the free energy of the products (the final state) minus the free energy of the reactants (the initial state).* The ΔG of a reaction is independent of the molecular mechanism of the transformation. For example, the ΔG for the oxidation of glucose to CO_2 and H_2O is the same whether it takes place by combustion or by a series of enzyme-catalyzed steps in a cell.

5. *The ΔG provides no information about the rate of a reaction.* A negative ΔG indicates that a reaction *can* take place spontaneously, but it does not signify whether it will proceed at a perceptible rate. As will be discussed shortly (Section 8.3), the rate of a reaction depends on the *free energy of activation* ($\Delta G^{\ddagger}$), which is largely unrelated to the ΔG of the reaction.

The standard free-energy change of a reaction is related to the equilibrium constant

As for any reaction, we need to be able to determine ΔG for an enzyme-catalyzed reaction to know whether the reaction is spontaneous or requires an input of energy. To determine this important thermodynamic parameter, we need to take into account the nature of both the reactants and the products as well as their concentrations.

Consider the reaction

$$A + B \rightleftharpoons C + D$$

The ΔG of this reaction is given by

$$\Delta G = \Delta G^{\circ} + RT \ln \frac{[C][D]}{[A][B]} \tag{1}$$

in which ΔG° is the *standard free-energy change,* R is the gas constant, T is the absolute temperature, and [A], [B], [C], and [D] are the molar concentrations (more precisely, the activities) of the reactants. ΔG° is the free-energy change for this reaction under standard conditions—that is, when each of the reactants A, B, C, and D is present at a concentration of 1.0 M (for a gas, the standard state is usually chosen to be 1 atmosphere). Thus, the ΔG of a reaction depends on the *nature* of the reactants (expressed in the ΔG° term of equation 1) and on their *concentrations* (expressed in the logarithmic term of equation 1).

A convention has been adopted to simplify free-energy calculations for biochemical reactions. The standard state is defined as having a pH of 7. Consequently, when H^+ is a reactant, its activity has the value 1 (corresponding to a pH of 7) in equations 1 and 3 (below). The activity of water also is taken to be 1 in these equations. The *standard free-energy change at pH 7,* denoted by the symbol $\Delta G^{\circ\prime}$, will be used throughout this book. The *kilojoule* (abbreviated kJ) and the *kilocalorie* (kcal) will be used as the units of energy. One kilojoule is equivalent to 0.239 kilocalorie.

A simple way to determine $\Delta G^{\circ\prime}$ is to measure the concentrations of reactants and products when the reaction has reached equilibrium. At equilibrium, there is no net change in reactants and products; in essence, the reaction has stopped and $\Delta G = 0$. At equilibrium, equation 1 then becomes

$$0 = \Delta G^{\circ\prime} + RT \ln \frac{[C][D]}{[A][B]} \tag{2}$$

and so

$$\Delta G^{\circ\prime} = -RT \ln \frac{[C][D]}{[A][B]} \tag{3}$$

The equilibrium constant under standard conditions, K'_{eq} is defined as

$$K'_{eq} = \frac{[C][D]}{[A][B]} \tag{4}$$

Substituting equation 4 into equation 3 gives

$$\Delta G^{\circ\prime} = -RT \ln K'_{eq} \tag{5}$$

which can be rearranged to give

$$K'_{eq} = e^{-\Delta G^{\circ\prime}/RT} \tag{6}$$

Units of energy

A *kilojoule* (kJ) is equal to 1000 J.
A *joule* (J) is the amount of energy needed to apply a 1-newton force over a distance of 1 meter.
A *kilocalorie* (kcal) is equal to 1000 cal.
A *calorie* (cal) is equivalent to the amount of heat required to raise the temperature of 1 gram of water from 14.5°C to 15.5°C.
1 kJ = 0.239 kcal.

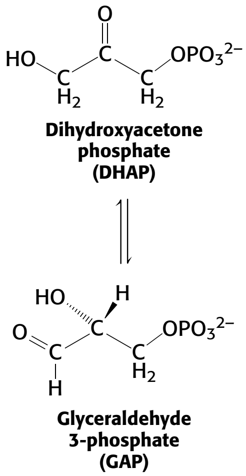
TABLE 8.3 Relation between $\Delta G^{\circ\prime}$ and K'_{eq} (at 25°C)

	$\Delta G^{\circ\prime}$	
K'_{eq}	kJ mol^{-1}	kcal mol^{-1}
10^{-5}	28.53	6.82
10^{-4}	22.84	5.46
10^{-3}	17.11	4.09
10^{-2}	11.42	2.73
10^{-1}	5.69	1.36
1	0.00	0.00
10	−5.69	−1.36
10^{2}	−11.42	−2.73
10^{3}	−17.11	−4.09
10^{4}	−22.84	−5.46
10^{5}	−28.53	−6.82

Substituting $R = 8.315 \times 10^{-3}$ kJ mol^{-1} deg^{-1} and $T = 298$ K (corresponding to 25° C) gives

$$K'_{\text{eq}} = e^{-\Delta G^{\circ\prime}/2.47} \tag{7}$$

where $\Delta G^{\circ\prime}$ is here expressed in kilojoules per mole because of the choice of the units for R in equation 7. Thus, the standard free energy and the equilibrium constant of a reaction are related by a simple expression. For example, an equilibrium constant of 10 gives a standard free-energy change of -5.69 kJ mol^{-1} (-1.36 kcal mol^{-1}) at 25° C (Table 8.3). Note that, for each 10-fold change in the equilibrium constant, the $\Delta G^{\circ\prime}$ changes by 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}).


Dihydroxyacetone phosphate (DHAP)
Glyceraldehyde 3-phosphate (GAP)

As an example, let us calculate $\Delta G^{\circ\prime}$ and ΔG for the isomerization of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP). This reaction takes place in glycolysis (Chapter 16). At equilibrium, the ratio of GAP to DHAP is 0.0475 at 25°C (298 K) and pH 7. Hence, $K'_{\text{eq}} = 0.0475$. The standard free-energy change for this reaction is then calculated from equation 5:

$$\begin{aligned}\Delta G^{\circ\prime} &= -RT \ln K'_{\text{eq}} \\ &= -8.315 \times 10^{-3} \times 298 \times \ln(0.0475) \\ &= +7.53 \text{ kJ mol}^{-1} (+1.80 \text{ kcal mol}^{-1})\end{aligned}$$

Under these conditions, the reaction is endergonic. DHAP will not spontaneously convert into GAP.

Now let us calculate ΔG for this reaction when the initial concentration of DHAP is 2×10^{-4} M and the initial concentration of GAP is 3×10^{-6} M. Substituting these values into equation 1 gives

$$\begin{aligned}\Delta G &= 7.53 \text{ kJ mol}^{-1} + RT \ln \frac{3 \times 10^{-6}\text{ M}}{2 \times 10^{-4}\text{ M}} \\ &= 7.53 \text{ kJ mol}^{-1} - 10.42 \text{ kJ mol}^{-1} \\ &= -2.89 \text{ kJ mol}^{-1} (-0.69 \text{ kcal mol}^{-1})\end{aligned}$$

This negative value for the ΔG indicates that the isomerization of DHAP to GAP is exergonic and can take place spontaneously when these species are present at the preceding concentrations. Note that ΔG for this reaction is negative, although $\Delta G^{\circ\prime}$ is positive. *It is important to stress that whether the ΔG for a reaction is larger, smaller, or the same as $\Delta G^{\circ\prime}$ depends on the concentrations of the reactants and products.* The criterion of spontaneity for a reaction is ΔG, not $\Delta G^{\circ\prime}$. This point is important because reactions that are not spontaneous based on $\Delta G^{\circ\prime}$ can be made spontaneous by adjusting the concentrations of reactants and products. This principle is the basis of the coupling of reactions to form metabolic pathways (Chapter 15).

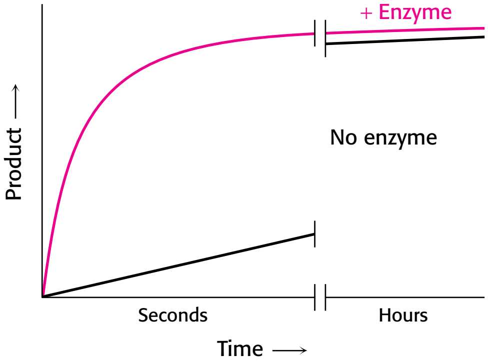


FIGURE 8.2 Enzymes accelerate the reaction rate. The same equilibrium point is reached but much more quickly in the presence of an enzyme.

Enzymes alter only the reaction rate and not the reaction equilibrium

Because enzymes are such superb catalysts, it is tempting to ascribe to them powers that they do not have. An enzyme cannot alter the laws of thermodynamics and *consequently cannot alter the equilibrium of a chemical reaction*. Consider an enzyme-catalyzed reaction, the conversion of substrate, S, into product, P. Figure 8.2 shows the rate of product formation with time in the presence and absence of enzyme. Note that the amount of product formed is the same whether or not the enzyme is present but, in the present example, the amount of product formed in seconds when the enzyme is present might take hours (or centuries, see Table 8.1) to form if the enzyme were absent.

Why does the rate of product formation level off with time? The reaction has reached equilibrium. Substrate S is still being converted into product P, but P is being converted into S at a rate such that the amount of P present stays the same.

Let us examine the equilibrium in a more quantitative way. Suppose that, in the absence of enzyme, the forward rate constant (k_F) for the conversion of S into P is $10^{-4}\ s^{-1}$ and the reverse rate constant (k_R) for the conversion of P into S is $10^{-6}\ s^{-1}$. The equilibrium constant K is given by the ratio of these rate constants:

$$\mathrm{S} \underset{10^{-6}\,\mathrm{s}^{-1}}{\overset{10^{-4}\,\mathrm{s}^{-1}}{\rightleftharpoons}} \mathrm{P}$$

$$K = \frac{[\mathrm{P}]}{[\mathrm{S}]} = \frac{k_F}{k_R} = \frac{10^{-4}}{10^{-6}} = 100$$

The equilibrium concentration of P is 100 times that of S, whether or not enzyme is present. However, it might take a very long time to approach this equilibrium without enzyme, whereas equilibrium would be attained rapidly in the presence of a suitable enzyme (Table 8.1). *Enzymes accelerate the attainment of equilibria but do not shift their positions. The equilibrium position is a function only of the free-energy difference between reactants and products.*

8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

The free-energy difference between reactants and products accounts for the equilibrium of the reaction, but enzymes accelerate how quickly this equilibrium is attained. How can we explain the rate enhancement in terms of thermodynamics? To do so, we have to consider not the end points of the reaction but the chemical pathway between the end points.

A chemical reaction of substrate S to form product P goes through a *transition state* $X^‡$ that has a higher free energy than does either S or P.

$$\mathrm{S} \longrightarrow \mathrm{X}^{‡} \longrightarrow \mathrm{P}$$

The double dagger denotes the transition state. The transition state is a transitory molecular structure that is no longer the substrate but is not yet the product. The transition state is the least-stable and most-seldom-occupied species along the reaction pathway because it is the one with the highest free energy. The difference in free energy between the transition state and the substrate is called the *Gibbs free energy of activation* or simply the *activation energy,* symbolized by $\Delta G^‡$ (Figure 8.3).

$$\Delta G^{‡} = G_X^{‡} - G_S$$

Note that the energy of activation, or $\Delta G^‡$, does not enter into the final ΔG calculation for the reaction, because the energy required to generate the transition state is released when the transition state forms the product. The activation-energy barrier immediately suggests how an enzyme enhances the reaction rate without altering ΔG of the reaction: enzymes function to lower the activation energy, or, in other words, *enzymes facilitate the formation of the transition state.*

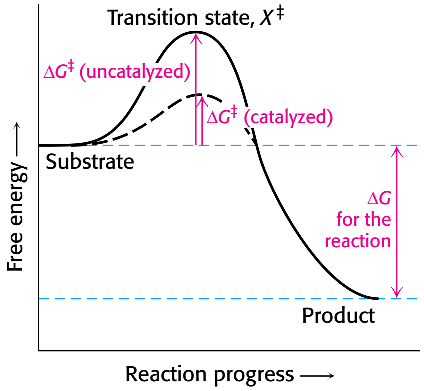


FIGURE 8.3 Enzymes decrease the activation energy. Enzymes accelerate reactions by decreasing $\Delta G^‡$, the free energy of activation.

One approach to understanding the increase in reaction rates achieved by enzymes is to assume that the transition state ($X^\ddagger$) and the substrate (S) are in equilibrium.

$$S \rightleftharpoons X^\ddagger \longrightarrow P$$

in which $K^\ddagger$ is the equilibrium constant for the formation of $X^\ddagger$ and v is the rate of formation of product from $X^\ddagger$. The rate of the reaction v is proportional to the concentration of $X^\ddagger$,

$$v \propto [X^\ddagger],$$

because only $X^\ddagger$ can be converted into product. The concentration of $X^\ddagger$ at equilibrium is in turn related to the free-energy difference $\Delta G^\ddagger$ between $X^\ddagger$ and S; the greater the difference in free energy between these two states, the smaller the amount of $X^\ddagger$. Thus, the overall rate of reaction V depends on $\Delta G^\ddagger$. Specifically,

$$V = v[X^\ddagger] = \frac{kT}{h}[S]e^{-\Delta G^\ddagger/RT}$$

In this equation, k is Boltzmann's constant, and h is Planck's constant. The value of kT/h at 25°C is $6.6 \times 10^{12}\ s^{-1}$. Suppose that the free energy of activation is 28.53 kJ mol^{-1} (6.82 kcal mol^{-1}). If we were to substitute this value of ΔG in equation 7 (as shown in Table 8.3), this free-energy difference would result when the ratio $[X^\ddagger]/[S]$ is 10^{-5}. If we assume for simplicity's sake that $[S] = 1$ M, then the reaction rate V is $6.2 \times 10^7\ s^{-1}$. If $\Delta G^\ddagger$ were lowered by 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}), the ratio $[X^\ddagger]/[S]$ would then be 10^{-4}, and the reaction rate would be $6.2 \times 10^8\ s^{-1}$. A decrease of 5.69 kJ mol^{-1} in $\Delta G^\ddagger$ yields a 10-fold larger V. A relatively small decrease in $\Delta G^\ddagger$ (20% in this particular reaction) results in a much greater increase in V.

Thus, we see the key to how enzymes operate: *enzymes accelerate reactions by decreasing* $\Delta G^\ddagger$, *the activation energy*. The combination of substrate and enzyme creates a reaction pathway whose transition-state energy is lower than that of the reaction in the absence of enzyme (Figure 8.3). Because the activation energy is lower, more molecules have the energy required to reach the transition state. Decreasing the activation barrier is analogous to lowering the height of a high-jump bar; more athletes will be able to clear the bar. *The essence of catalysis is facilitating the formation of the transition state.*

> "I think that enzymes are molecules that are complementary in structure to the activated complexes of the reactions that they catalyze, that is, to the molecular configuration that is intermediate between the reacting substances and the products of reaction for these catalyzed processes. The attraction of the enzyme molecule for the activated complex would thus lead to a decrease in its energy and hence to a decrease in the energy of activation of the reaction and to an increase in the rate of reaction."
>
> —Linus Pauling
> *Nature* 161:707, 1948

The formation of an enzyme–substrate complex is the first step in enzymatic catalysis

Much of the catalytic power of enzymes comes from their binding to and then altering the structure of the substrate to promote the formation of the transition state. Thus, the first step in catalysis is the formation of an *enzyme–substrate* (ES) complex. Substrates bind to a specific region of the enzyme called the *active site*. Most enzymes are highly selective in the substrates that they bind. Indeed, the catalytic specificity of enzymes depends in part on the specificity of binding.

What is the evidence for the existence of an enzyme–substrate complex?

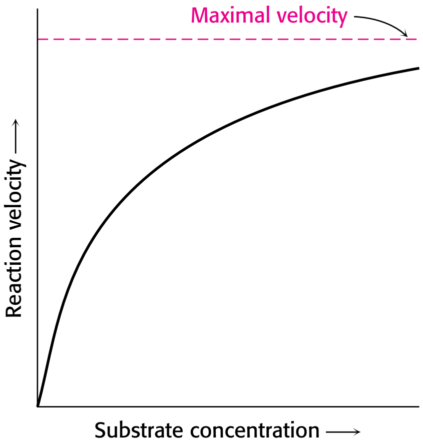


FIGURE 8.4 Reaction velocity versus substrate concentration in an enzyme-catalyzed reaction. An enzyme-catalyzed reaction approaches a maximal velocity.

1. The first clue was the observation that, at a constant concentration of enzyme, the reaction rate increases with increasing substrate concentration until a maximal velocity is reached (Figure 8.4). In contrast, uncatalyzed reactions do not show this saturation effect. *The fact that an enzyme-catalyzed reaction has a maximal velocity suggests the formation of a discrete ES complex.* At a sufficiently high substrate concentration, all the catalytic sites are filled, or saturated, and so the reaction rate cannot increase. Although indirect, the

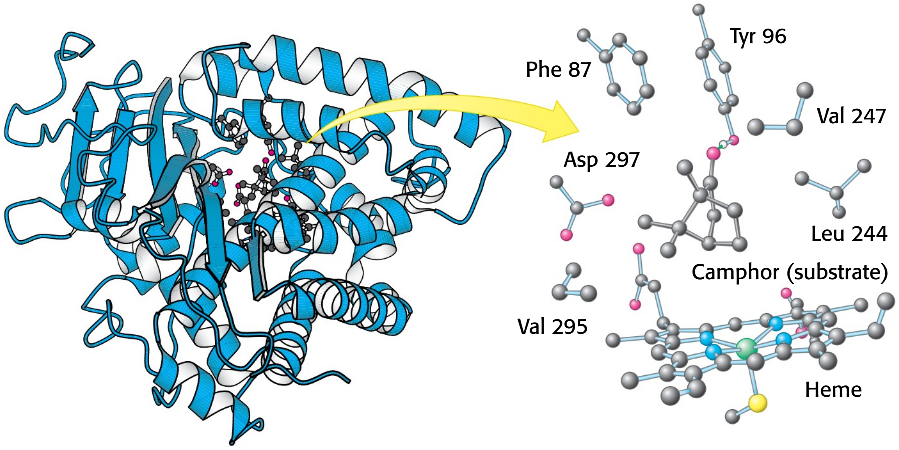


FIGURE 8.5 Structure of an enzyme–substrate complex. (Left) The enzyme cytochrome P450 is illustrated bound to its substrate camphor. (Right) *Notice* that, in the active site, the substrate is surrounded by residues from the enzyme. Note also the presence of a heme cofactor. [Drawn from 2CPP.pdb.]

ability to saturate an enzyme with substrate is the most general evidence for the existence of ES complexes.

2. The *spectroscopic characteristics* of many enzymes and substrates change on the formation of an ES complex. These changes are particularly striking if the enzyme contains a colored prosthetic group (Problem 39).

3. *X-ray crystallography* has provided high-resolution images of substrates and substrate analogs bound to the active sites of many enzymes (Figure 8.5). In Chapter 9, we will take a close look at several of these complexes.

The active sites of enzymes have some common features

The active site of an enzyme is the region that binds the substrates (and the cofactor, if any). It also contains the amino acid residues that directly participate in the making and breaking of bonds. These residues are called the *catalytic groups*. In essence, *the interaction of the enzyme and substrate at the active site promotes the formation of the transition state*. The active site is the region of the enzyme that most directly lowers the $\Delta G^{\ddagger}$ of the reaction, thus providing the rate-enhancement characteristic of enzyme action. Recall from Chapter 2 that proteins are not rigid structures, but are flexible and exist in an array of conformations. Thus, the interaction of the enzyme and substrate at the active site and the formation of the transition state is a dynamic process. Although enzymes differ widely in structure, specificity, and mode of catalysis, a number of generalizations concerning their active sites can be stated:

1. *The active site is a three-dimensional cleft, or crevice,* formed by groups that come from different parts of the amino acid sequence: indeed, residues far apart in the amino acid sequence may interact more strongly than adjacent residues in the sequence, which may be sterically constrained from interacting with one another. In lysozyme, an enzyme that degrades the cell walls of some bacteria, the important groups in the active site are contributed by residues numbered 35, 52, 62, 63, 101, and 108 in the sequence of 129 amino acids (Figure 8.6).

2. *The active site takes up a small part of the total volume of an enzyme*. Although most of the amino acid residues in an enzyme are not in contact with the substrate, the cooperative motions of the entire enzyme help to correctly position the catalytic residues at the active site. Experimental attempts to reduce the size of a catalytically active enzyme show that the minimum size requires about 100 amino acid

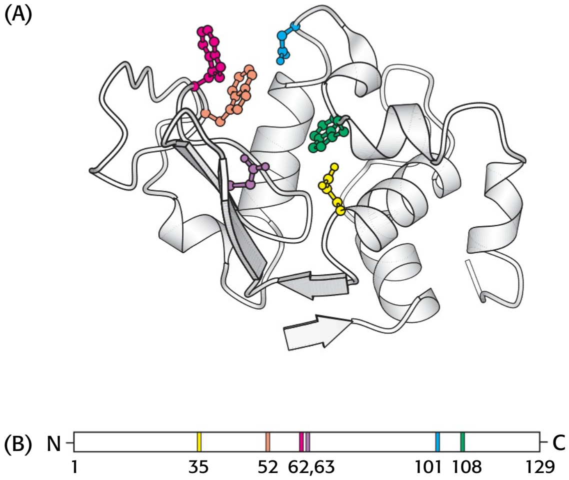


FIGURE 8.6 Active sites may include distant residues. (A) Ribbon diagram of the enzyme lysozyme with several components of the active site shown in color. (B) A schematic representation of the primary structure of lysozyme shows that the active site is composed of residues that come from different parts of the polypeptide chain. [Drawn from 6LYZ.pdb.]

residues. In fact, nearly all enzymes are made up of more than 100 amino acid residues, which gives them a mass greater than 10 kDa and a diameter of more than 25 Å, suggesting that all amino acids in the protein, not just those at the active site, are ultimately required to form a functional enzyme.

3. *Active sites are unique microenvironments*. In all enzymes of known structure, active sites are shaped like a cleft, or crevice, to which the substrates bind. Water is usually excluded unless it is a reactant. The nonpolar microenvironment of the cleft enhances the binding of substrates as well as catalysis. Nevertheless, the cleft may also contain polar residues, some of which may acquire special properties essential for substrate binding or catalysis. The internal positions of these polar residues are biologically crucial exceptions to the general rule that polar residues are located on the surface of proteins, exposed to water.

4. *Substrates are bound to enzymes by multiple weak attractions*. The noncovalent interactions in ES complexes are much weaker than covalent bonds, which have energies between -210 and -460 kJ mol^{-1} (between -50 and -110 kcal mol^{-1}). In contrast, ES complexes usually have equilibrium constants that range from 10^{-2} to 10^{-8} M, corresponding to free energies of interaction ranging from about -13 to -50 kJ mol^{-1} (from -3 to -12 kcal mol^{-1}). As discussed in Section 1.3, these weak reversible contacts are mediated by electrostatic interactions, hydrogen bonds, and van der Waals forces. Van der Waals forces become significant in binding only when numerous substrate atoms simultaneously come close to many enzyme atoms through the hydrophobic effect. Hence, the enzyme and substrate should have complementary shapes. The directional character of hydrogen bonds between enzyme and substrate often enforces a high degree of specificity, as seen in the RNA-degrading enzyme ribonuclease (Figure 8.7).

5. *The specificity of binding depends on the precisely defined arrangement of atoms in an active site*. Because the enzyme and the substrate interact by means of short-range forces that require close contact, a substrate must have a matching shape to fit into the site. Emil Fischer proposed the lock-and-key analogy in 1890 (Figure 8.8), which was the model for enzyme–substrate interaction for several decades. We now know that enzymes are flexible and that the shapes of the active sites can be markedly modified by the binding of substrate, a process of dynamic recognition called *induced fit* (Figure 8.9). Moreover, the substrate may bind to only certain conformations of the enzyme, in what is called *conformation selection*. Thus, the mechanism of catalysis is dynamic, involving structural changes with multiple intermediates of both reactants and the enzyme.

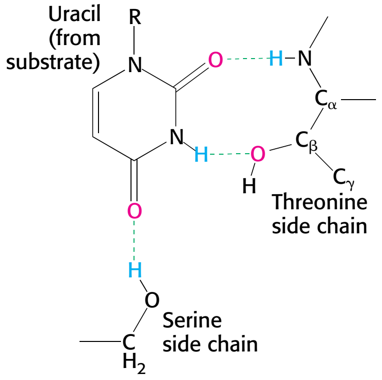


FIGURE 8.7 Hydrogen bonds between an enzyme and substrate. The enzyme ribonuclease forms hydrogen bonds with the uridine component of the substrate. [Information from F. M. Richards, H. W. Wyckoff, and N. Allewell. In *The Neurosciences: Second Study Program*, F. O. Schmidt, Ed. (Rockefeller University Press, 1970), p. 970.]

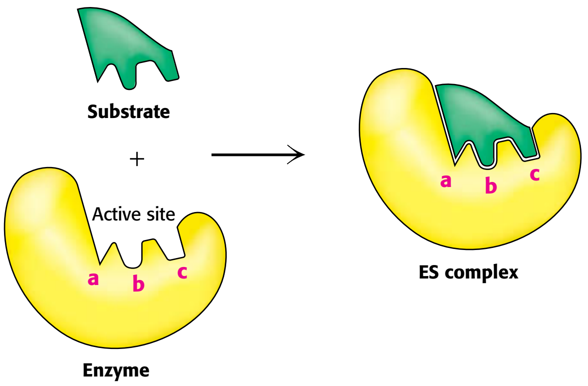


FIGURE 8.8 Lock-and-key model of enzyme–substrate binding. In this model, the active site of the unbound enzyme is complementary in shape to the substrate.

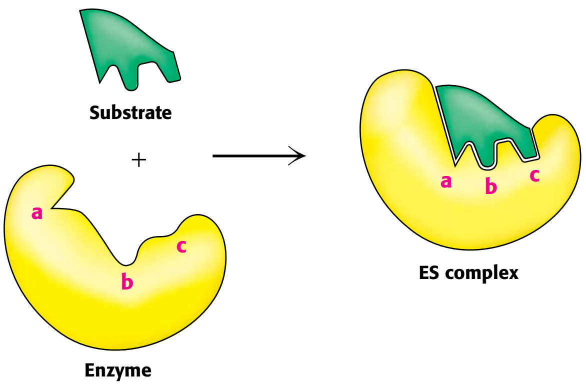


FIGURE 8.9 Induced-fit model of enzyme–substrate binding. In this model, the enzyme changes shape on substrate binding. The active site forms a shape complementary to the substrate only after the substrate has been bound.

The binding energy between enzyme and substrate is important for catalysis

Enzymes lower the activation energy, but where does the energy to lower the activation energy come from? Free energy is released by the formation of a large number of weak interactions between a complementary enzyme and its substrate. The free energy released on binding is called the *binding energy*. Only the correct substrate can participate in most or all of the interactions with the enzyme and thus maximize binding energy, accounting for the exquisite substrate specificity exhibited by many enzymes. Furthermore, *the full complement of such interactions is formed only when the substrate is converted into the transition state*. Thus, the maximal binding energy is released when the enzyme facilitates the formation of the transition state. The energy released by the interaction between the enzyme and the substrate can be thought of as lowering the activation energy. The interaction of the enzyme with the substrate and reaction intermediates is fleeting, with molecular movements resulting in the optimal alignment of functional groups at the active site so that maximum binding energy occurs only between the enzyme and the transition state, the least-stable reaction intermediate. However, the transition state is too unstable to exist for long. It collapses to either substrate or product, but which of the two accumulates is determined only by the energy difference between the substrate and the product—that is, by the ΔG of the reaction.

8.4 The Michaelis–Menten Model Accounts for the Kinetic Properties of Many Enzymes

The study of the rates of chemical reactions is called *kinetics,* and the study of the rates of enzyme-catalyzed reactions is called *enzyme kinetics*. A kinetic description of enzyme activity will help us understand how enzymes function. We begin by briefly examining some of the basic principles of reaction kinetics.

Kinetics is the study of reaction rates

What do we mean when we say the "rate" of a chemical reaction? Consider a simple reaction:

$$\mathrm{A} \longrightarrow \mathrm{P}$$

The rate V is the quantity of A that disappears in a specified unit of time. It is equal to the rate of the appearance of P, or the quantity of P that appears in a specified unit of time.

$$V = -\Delta\mathrm{A}/\Delta T = \Delta\mathrm{P}/\Delta T \tag{8}$$

If A is yellow and P is colorless, we can follow the decrease in the concentration of A by measuring the decrease in the intensity of yellow color with time. Consider only the change in the concentration of A for now. The rate of the reaction is directly related to the concentration of A by a proportionality constant, k, called the *rate constant*.

$$V = k[\mathrm{A}] \tag{9}$$

Reactions that are directly proportional to the reactant concentration are called *first-order reactions*. First-order rate constants have the units of s^{-1}.

Many important biochemical reactions include two reactants. For example,

$$2\mathrm{A} \longrightarrow \mathrm{P}$$

or

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{P}$$

They are called *bimolecular reactions* and the corresponding rate equations often take the form

$$V = k[\mathrm{A}]^2 \tag{10}$$

and

$$V = k[\mathrm{A}][\mathrm{B}] \tag{11}$$

The rate constants, called second-order rate constants, have the units $\mathrm{M}^{-1}\,\mathrm{s}^{-1}$.

Sometimes, second-order reactions can appear to be first-order reactions. For instance, in reaction 11, if B is present in excess and A is present at low concentrations, the reaction rate will be first order with respect to A and will not appear to depend on the concentration of B. These reactions are called *pseudo-first-order reactions,* and we will see them a number of times in our study of biochemistry.

Interestingly enough, under some conditions, a reaction can be zero order. In these cases, the rate is independent of reactant concentrations. Enzyme-catalyzed reactions can approximate zero-order reactions under some circumstances (p. 228).

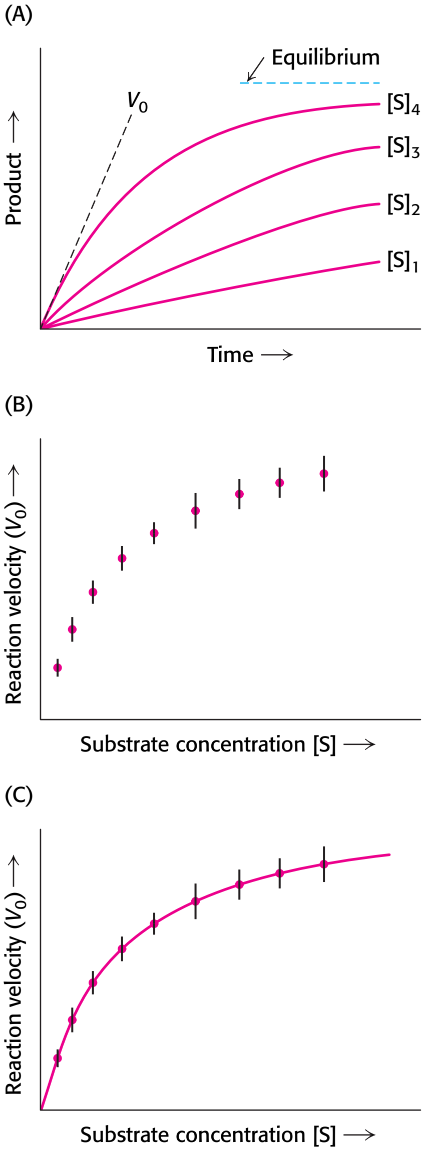


FIGURE 8.10 Determining the relation between initial velocity and substrate concentration. (A) The amount of product formed at different substrate concentrations is plotted as a function of time. The initial velocity (V_0) for each substrate concentration is determined from the slope of the curve at the beginning of a reaction, when the reverse reaction is insignificant. (B) The values for initial velocity determined in part A are then plotted, with error bars, against substrate concentration. (C) The data points are connected to clearly reveal the relationship of initial velocity to substrate concentration.

The steady-state assumption facilitates a description of enzyme kinetics

The simplest way to investigate the reaction rate is to follow the increase in reaction product as a function of time. First, the extent of product formation is determined as a function of time for a series of substrate concentrations (Figure 8.10A). As expected, in each case, the amount of product formed increases with time, although eventually a time is reached when there is *no net change* in the concentration of S or P. The enzyme is still actively converting substrate into product and vice versa, but the reaction equilibrium has been attained. However, enzyme kinetics is more readily comprehended if we consider only the forward reaction. We can define the rate of catalysis V_0, or the initial rate of catalysis, as the number of moles of product formed per second when the reaction is just beginning—that is, when $t \approx 0$ (Figure 8.10A). These experiments are repeated three to five times with each substrate concentration to insure the accuracy of and assess the variability of the values attained. Next, we plot V_0 versus the substrate concentration [S], assuming a constant amount of enzyme, showing the data points with error bars (Figure 8.10B). Finally, the data points are connected, yielding the results shown in Figure 8.10C. The rate of catalysis rises linearly as substrate concentration increases and then begins to level off and approach a maximum at higher substrate concentrations. For convenience, we will show idealized data without error bars, throughout the text, but it is important to keep in mind that in reality, all experiments are repeated multiple times.

In 1913, Leonor Michaelis and Maud Menten proposed a simple model to account for these kinetic characteristics. The critical feature in their treatment is that a specific ES complex is a necessary intermediate in catalysis. The model proposed is

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P}$$

An enzyme E combines with substrate S to form an ES complex, with a rate constant k_1. The ES complex has two possible fates. It can dissociate to E and S, with a rate constant k_{-1}, or it can proceed to form product P, with a rate constant k_2. The ES complex can also be reformed from E and P by the reverse reaction with a rate constant k_{-2}. However, as before, we can simplify these reactions by considering the rate of reaction at times close to zero (hence, V_0) when there is negligible product formation and thus no back reaction (k_{-2} [E][P] ≈ 0).

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{12}$$

Thus, for the graph in Figure 8.11, V_0 is determined for each substrate concentration by measuring the rate of product formation at early times before P accumulates (Figure 8.10A).

We want an expression that relates the rate of catalysis to the concentrations of substrate and enzyme and the rates of the individual steps. Our starting point is that the catalytic rate is equal to the product of the concentration of the ES complex and k_2.

$$V_0 = k_2[\mathrm{ES}] \tag{13}$$

Now we need to express [ES] in terms of known quantities. The rates of formation and breakdown of ES are given by

$$\text{Rate of formation of ES} = k_1[\mathrm{E}][\mathrm{S}] \tag{14}$$

$$\text{Rate of breakdown of ES} = (k_{-1} + k_2)[\mathrm{ES}] \tag{15}$$

We will use the *steady-state assumption* to simplify matters. In a steady state, the concentrations of intermediates—in this case, [ES]—stay the same even if the concentrations of starting materials and products are changing. This steady state is reached when the rates of formation and breakdown of the ES complex are equal. Setting the right-hand sides of equations 14 and 15 equal gives

$$k_1[\mathrm{E}][\mathrm{S}] = (k_{-1} + k_2)[\mathrm{ES}] \tag{16}$$

By rearranging equation 16, we obtain

$$[\mathrm{E}][\mathrm{S}]/[\mathrm{ES}] = (k_{-1} + k_2)/k_1 \tag{17}$$

Equation 17 can be simplified by defining a new constant, K_M, called the Michaelis constant:

$$K_M = \frac{k_{-1} + k_2}{k_1} \tag{18}$$

Note that K_M has the units of concentration and is independent of enzyme and substrate concentrations. As will be explained, K_M is an important characteristic of enzyme–substrate interactions.

Inserting equation 18 into equation 17 and solving for [ES] yields

$$[\mathrm{ES}] = \frac{[\mathrm{E}][\mathrm{S}]}{K_M} \tag{19}$$

Now let us examine the numerator of equation 19. Because the substrate is usually present at a much higher concentration than that of the enzyme, the concentration of uncombined substrate [S] is very nearly equal to the total substrate concentration. The concentration of uncombined enzyme [E] is equal to the total enzyme concentration $[\mathrm{E}]_T$ minus the concentration of the ES complex:

$$[\mathrm{E}] = [\mathrm{E}]_T - [\mathrm{ES}] \tag{20}$$

Substituting this expression for [E] in equation 19 gives

$$[\mathrm{ES}] = \frac{([\mathrm{E}]_T - [\mathrm{ES}])[\mathrm{S}]}{K_M} \tag{21}$$

Solving equation 21 for [ES] gives

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_T[\mathrm{S}]/K_M}{1 + [\mathrm{S}]/K_M} \tag{22}$$

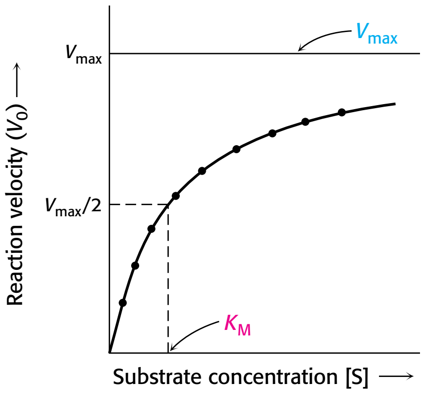


FIGURE 8.11 Michaelis–Menten kinetics. A plot of the reaction velocity (V_0) as a function of the substrate concentration [S] for an enzyme that obeys Michaelis–Menten kinetics shows that the maximal velocity (V_{max}) is approached asymptotically. The Michaelis constant (K_M) is the substrate concentration yielding a velocity of $V_{max}/2$.

or

$$[\mathrm{ES}] = [\mathrm{E}]_\mathrm{T}\,\frac{[\mathrm{S}]}{[\mathrm{S}] + K_\mathrm{M}} \tag{23}$$

By substituting this expression for [ES] into equation 13, we obtain

$$V_0 = k_2[\mathrm{E}]_\mathrm{T}\,\frac{[\mathrm{S}]}{[\mathrm{S}] + K_\mathrm{M}} \tag{24}$$

The *maximal rate,* V_max, is attained when the catalytic sites on the enzyme are saturated with substrate—that is, when $[\mathrm{ES}] = [\mathrm{E}]_\mathrm{T}$. Thus,

$$V_\mathrm{max} = k_2[\mathrm{E}]_\mathrm{T} \tag{25}$$

Substituting equation 25 into equation 24 yields the *Michaelis–Menten equation:*

$$V_0 = V_\mathrm{max}\frac{[\mathrm{S}]}{[\mathrm{S}] + K_\mathrm{M}} \tag{26}$$

This equation accounts for the kinetic data given in Figure 8.11. At very low substrate concentration, when [S] is much less than K_M, $V_0 = (V_\mathrm{max}/K_\mathrm{M})[\mathrm{S}]$; that is, the reaction is first order with the rate directly proportional to the substrate concentration. At high substrate concentration, when [S] is much greater than K_M, $V_0 = V_\mathrm{max}$; that is, the rate is maximal. The reaction is zero order, independent of substrate concentration.

The significance of K_M is clear when we set $[\mathrm{S}] = K_\mathrm{M}$ in equation 26. When $[\mathrm{S}] = K_\mathrm{M}$, then $V_0 = V_\mathrm{max}/2$. Thus, *K_M is equal to the substrate concentration at which the reaction rate is half its maximal value*. As we will see, K_M is an important characteristic of an enzyme-catalyzed reaction and is significant for its biological function.

Variations in K_M can have physiological consequences

The physiological consequence of K_M is illustrated by the sensitivity of some persons to ethanol. Such persons exhibit facial flushing and rapid heart rate (tachycardia) after ingesting even small amounts of alcohol. In the liver, alcohol dehydrogenase converts ethanol into acetaldehyde.

$$\underset{\text{Ethanol}}{\mathrm{CH_3CH_2OH}} + \mathrm{NAD^+} \xrightleftharpoons[]{\substack{\text{Alcohol}\\ \text{dehydrogenase}}} \underset{\text{Acetaldehyde}}{\mathrm{CH_3CHO}} + \mathrm{NADH} + \mathrm{H^+}$$

Normally, the acetaldehyde, which is the cause of the symptoms when present at high concentrations, is processed to acetate by aldehyde dehydrogenase.

$$\mathrm{CH_3CHO} + \mathrm{NAD^+} + \mathrm{H_2O} \xrightleftharpoons[]{\substack{\text{Aldehyde}\\ \text{dehydrogenase}}} \underset{\text{Acetate}}{\mathrm{CH_3COO^-}} + \mathrm{NADH} + 2\mathrm{H^+}$$

Most people have two forms of the aldehyde dehydrogenase, a low K_M mitochondrial form and a high K_M cytoplasmic form. In susceptible persons, the mitochondrial enzyme is less active owing to the substitution of a single amino acid, and acetaldehyde is processed only by the cytoplasmic enzyme. Because this enzyme has a high K_M, it achieves a high rate of catalysis only at very high concentrations of acetaldehyde. Consequently, less acetaldehyde is converted into acetate; excess acetaldehyde escapes into the blood and accounts for the physiological effects.

K_M and V_max values can be determined by several means

K_M is equal to the substrate concentration that yields $V_\mathrm{max}/2$; however V_max, like perfection, is only approached but never attained. How, then, can we experimentally determine K_M and V_max, and how do these parameters enhance our understanding of enzyme-catalyzed reactions? The Michaelis constant, K_M, and the maximal rate, V_{max}, can be readily derived from rates of catalysis measured at a variety of substrate concentrations if an enzyme operates according to the simple scheme given in equation 26. The derivation of K_M and V_{max} is most commonly achieved with the use of curve-fitting programs on a computer. However, an older method, although rarely used because the data points at high and low concentrations are weighted differently and thus sensitive to errors, is a source of further insight into the meaning of K_M and V_{max}.

Before the availability of computers, the determination of K_M and V_{max} values required algebraic manipulation of the Michaelis–Menten equation. The Michaelis–Menten equation is transformed into one that gives a straight-line plot that yields values for V_{max} and K_M. Taking the reciprocal of both sides of equation 26 gives

$$\frac{1}{V_0} = \frac{K_M}{V_{max}} \cdot \frac{1}{S} + \frac{1}{V_{max}} \tag{27}$$

A plot of $1/V_0$ versus $1/[S]$, called a *Lineweaver–Burk* or *double-reciprocal plot,* yields a straight line with a y-intercept of $1/V_{max}$ and a slope of K_M/V_{max} (Figure 8.12). The intercept on the x-axis is $-1/K_M$.

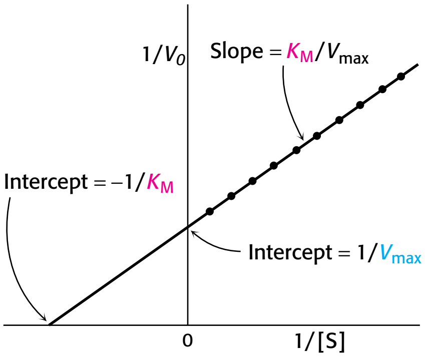


FIGURE 8.12 A double-reciprocal or Lineweaver–Burk plot. A double-reciprocal plot of enzyme kinetics is generated by plotting $1/V_0$ as a function of $1/[S]$. The slope is K_M/V_{max}, the intercept on the vertical axis is $1/V_{max}$, and the intercept on the horizontal axis is $-1/K_M$.

K_M and V_{max} values are important enzyme characteristics

The K_M values of enzymes range widely (Table 8.4). For most enzymes, K_M lies between 10^{-1} and 10^{-7} M. The K_M value for an enzyme depends on the particular substrate and on environmental conditions such as pH, temperature, and ionic strength. The Michaelis constant, K_M, as already noted, is equal to the concentration of substrate at which half the active sites are filled. Thus, K_M provides a measure of the substrate concentration required for significant catalysis to take place. For many enzymes, experimental evidence suggests that the K_M value provides an approximation of the substrate concentration in vivo, which in turn suggests that most enzymes evolved to have a K_M approximately equal to the substrate concentration commonly available. Why might it be beneficial to have a K_M value approximately equal to the commonly available substrate concentration? If the normal concentration of substrate is near K_M, the enzyme will display significant activity and yet the activity will be sensitive to changes in environmental conditions—that is, changes in substrate concentration. At values below K_M, enzymes are very sensitive to changes in substrate concentration but display little activity. At substrate values well above K_M, enzymes have great catalytic activity but are insensitive to changes in substrate concentration. Thus, with the normal substrate concentration being approximately K_M, the enzymes have significant activity ($1/2\ V_{max}$) but are still sensitive to changes in substrate concentration.

Under certain circumstances, K_M reflects the strength of the enzyme–substrate interaction. In equation 18, K_M is defined as $(k_{-1} + k_2)/k_1$. Consider a case in which k_{-1} is much greater than k_2. Under such circumstances, the ES complex dissociates to E and S much more rapidly than product is formed. Under these conditions ($k_{-1} \gg k_2$)

$$K_M \approx \frac{k_{-1}}{k_1} \tag{28}$$

Equation 28 describes the *dissociation constant* of the ES complex.

$$K_{ES} = \frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1} \tag{29}$$

TABLE 8.4 K_M values of some enzymes

Enzyme	Substrate	K_M (μM)
Chymotrypsin	Acetyl-L-tryptophanamide	5000
Lysozyme	Hexa-*N*-acetylglucosamine	6
β-Galactosidase	Lactose	4000
Threonine deaminase	Threonine	5000
Carbonic anhydrase	CO_2	8000
Penicillinase	Benzylpenicillin	50
Pyruvate carboxylase	Pyruvate	400
	HCO_3^-	1000
	ATP	60
Arginine-tRNA synthetase	Arginine	3
	tRNA	0.4
	ATP	300

In other words, K_M is equal to the dissociation constant of the ES complex if k_2 is much smaller *than* k_{-1}. When this condition is met, K_M is a measure of the strength of the ES complex: a high K_M indicates weak binding; a low K_M indicates strong binding. It must be stressed that K_M indicates the affinity of the ES complex only when k_{-1} is much greater than k_2.

The maximal rate, V_{max}, reveals the *turnover number* of an enzyme, which is the number of substrate molecules converted into product by an enzyme molecule in a unit time when the enzyme is fully saturated with substrate. It is equal to the rate constant k_2, which is also called k_{cat}. The maximal rate, V_{max}, reveals the turnover number of an enzyme if the concentration of active sites $[E]_T$ is known, because

$$V_{max} = k_{cat}[E]_T \tag{30}$$

and thus

$$k_{cat} = V_{max}/[E]_T \tag{31}$$

For example, a 10^{-6} M solution of carbonic anhydrase catalyzes the formation of 0.6 M H_2CO_3 per second when the enzyme is fully saturated with substrate. Hence, k_{cat} is $6 \times 10^5\ s^{-1}$. This turnover number is one of the largest known. Each catalyzed reaction takes place in a time equal to, on average, $1/k_{cat}$, which is 1.7 μs for carbonic anhydrase. The turnover numbers of most enzymes with their physiological substrates range from 1 to 10^4 per second (Table 8.5).

TABLE 8.5 Turnover numbers of some enzymes

Enzyme	Turnover number (per second)
Carbonic anhydrase	600,000
3-Ketosteroid isomerase	280,000
Acetylcholinesterase	25,000
Penicillinase	2,000
Lactate dehydrogenase	1,000
Chymotrypsin	100
DNA polymerase I	15
Tryptophan synthetase	2
Lysozyme	0.5

K_M and V_{max} also permit the determination of f_{ES}, the fraction of active sites filled. This relation of f_{ES} to K_M and V_{max} is given by the following equation:

$$f_{ES} = \frac{V}{V_{max}} = \frac{[S]}{[S] + K_M} \tag{32}$$

k_{cat}/K_M is a measure of catalytic efficiency

When the substrate concentration is much greater than K_M, the rate of catalysis is equal to V_{max}, which is a function of k_{cat}, the turnover number, as already described. However, most enzymes are not normally saturated with substrate. Under physiological conditions, the $[S]/K_M$ ratio is typically between 0.01 and 1.0. When $[S] \ll K_M$, the enzymatic rate is much less than k_{cat} because most of the active sites are unoccupied. Is there a number that characterizes the kinetics of an enzyme under these more typical cellular conditions? Indeed there is, as can be shown by combining equations 13 and 19 to give

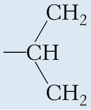
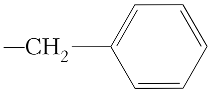
$$V_0 = \frac{k_{cat}}{K_M}[E][S] \tag{33}$$

When $[S] \ll K_M$, the concentration of free enzyme [E], is nearly equal to the total concentration of enzyme $[E]_T$; so

$$V_0 = \frac{k_{cat}}{K_M}[S][E]_T \tag{34}$$

Thus, when $[S] \ll K_M$, the enzymatic velocity depends on the values of k_{cat}/K_M, [S], and $[E]_T$. Under these conditions, k_{cat}/K_M is the rate constant for the interaction of S and E. The rate constant k_{cat}/K_M, called the *specificity constant*, is a measure of catalytic efficiency because it takes into account both the rate of catalysis with a particular substrate (k_{cat}) and the nature of the enzyme–substrate interaction (K_M). For instance, by using k_{cat}/K_M values, we can compare an enzyme's preference for different substrates. Table 8.6 shows the k_{cat}/K_M values for several different substrates of chymotrypsin. Chymotrypsin clearly has a preference for cleaving next to bulky, hydrophobic side chains.

TABLE 8.6 Substrate preferences of chymotrypsin

Amino acid in ester	Amino acid side chain	k_{cat}/K_M (s^{-1} M^{-1})
Glycine	$—H$	1.3×10^{-1}
Valine	 $-CH(CH_2)_2$	2.0
Norvaline	$—CH_2CH_2CH_3$	3.6×10^2
Norleucine	$—CH_2CH_2CH_2CH_3$	3.0×10^3
Phenylalanine		1.0×10^5

Source: Information from A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999), Table 7.3.

How efficient can an enzyme be? We can approach this question by determining whether there are any physical limits on the value of k_{cat}/K_M. Note that the k_{cat} K_M *ratio* depends on k_1, k_{-1}, and k_{cat}, as can be shown by substituting for K_M.

$$k_{cat}/K_M = \frac{k_{cat}k_1}{k_{-1} + k_{cat}} = \left(\frac{k_{cat}}{k_{-1} + k_{cat}}\right)k_1 < k_1 \qquad (35)$$

Suppose that the rate of formation of product (k_{cat}) is much faster than the rate of dissociation of the ES complex (k_{-1}). The value of k_{cat}/K_M then approaches k_1. Thus, the ultimate limit on the value of k_{cat}/K_M is set by k_1, the rate of formation of the ES complex. *This rate cannot be faster than the diffusion-controlled encounter of an enzyme and its substrate.* Diffusion limits the value of k_1 and so it cannot be higher than between 10^8 and 10^9 s^{-1} M^{-1}. Hence, the upper limit on k_{cat}/K_M is between 10^8 and 10^9 s^{-1} M^{-1}.

The k_{cat}/K_M ratios of the enzymes superoxide dismutase, acetylcholinesterase, and triose phosphate isomerase are between 10^8 and 10^9 s^{-1} M^{-1}. Enzymes that have k_{cat}/K_M ratios at the upper limits have attained *kinetic perfection. Their catalytic velocity is restricted only by the rate at which they encounter substrate in the solution* (Table 8.7). Any further gain in catalytic rate can come only by decreasing the time for diffusion of the substrate into the enzyme's immediate environment. Remember that the active site is only a small part of the total enzyme structure. Yet, for catalytically perfect enzymes, every encounter between enzyme and substrate is productive. In these cases, there may be attractive electrostatic forces on the enzyme that entice the substrate to the active site. These forces are sometimes referred to poetically as *Circe effects*.

TABLE 8.7 Enzymes for which k_{cat}/K_M is close to the diffusion-controlled rate of encounter

Enzyme	k_{cat}/K_M (s^{-1} M^{-1})
Acetylcholinesterase	1.6×10^8
Carbonic anhydrase	8.3×10^7
Catalase	4×10^7
Crotonase	2.8×10^8
Fumarase	1.6×10^8
Triose phosphate isomerase	2.4×10^8
β-Lactamase	1×10^8
Superoxide dismutase	7×10^9

Source: Information from A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999), Table 4.5.

Circe effect

The utilization of attractive forces to lure a substrate into a site in which it undergoes a transformation of structure, as defined by William P. Jencks, an enzymologist, who coined the term.

A goddess of Greek mythology, Circe lured Odysseus's men to her house and then transformed them into pigs.

The diffusion of a substrate throughout a solution can also be partly overcome by confining substrates and products in the limited volume of a multienzyme complex. Indeed, some series of enzymes are organized into complexes so that the product of one enzyme is very rapidly found by the next enzyme. In effect, products are channeled from one enzyme to the next, much as in an assembly line.

Most biochemical reactions include multiple substrates

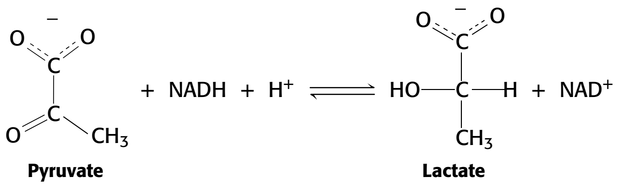
Most reactions in biological systems start with two substrates and yield two products. They can be represented by the bisubstrate reaction:

$$A + B \rightleftharpoons P + Q$$

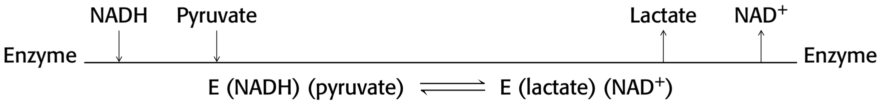
Many such reactions transfer a functional group, such as a phosphoryl or an ammonium group, from one substrate to the other. Those that are oxidation–reduction reactions transfer electrons between substrates. Multiple substrate reactions can be divided into two classes: *sequential* reactions and *double-displacement* reactions.

Sequential reactions. In *sequential reactions,* all substrates must bind to the enzyme before any product is released. Consequently, in a bisubstrate reaction, a *ternary complex* of the enzyme and both substrates forms. Sequential mechanisms are of two types: ordered, in which the substrates bind the enzyme in a defined sequence, and random.

Many enzymes that have NAD$^+$ or NADH as a substrate exhibit the ordered sequential mechanism. Consider lactate dehydrogenase, an important enzyme in glucose metabolism (Section 16.1). This enzyme reduces pyruvate to lactate while oxidizing NADH to NAD$^+$.

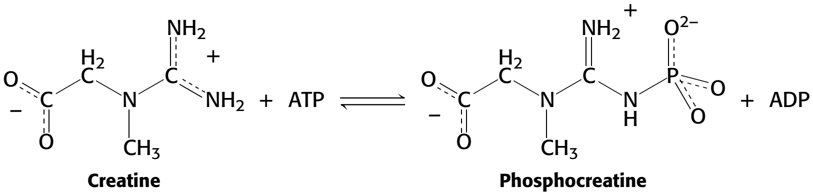

+ NADH + H+
HO—C—H + NAD+
CH3
Pyruvate
Lactate

In the ordered sequential mechanism, the coenzyme always binds first and the lactate is always released first. This sequence can be represented by using a notation developed by W. Wallace Cleland:

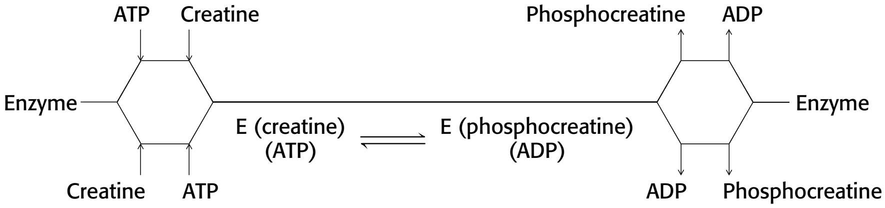

NADH
Pyruvate
Lactate
NAD+
Enzyme
Enzyme
E (NADH) (pyruvate)
E (lactate) (NAD+)

The enzyme exists as a ternary complex consisting of, first, the enzyme and substrates and, after catalysis, the enzyme and products.

In the random sequential mechanism, the order of the addition of substrates and the release of products is random. An example of a random sequential reaction is the formation of phosphocreatine and ADP from creatine and ATP which is catalyzed by creatine kinase (Section 15.2).

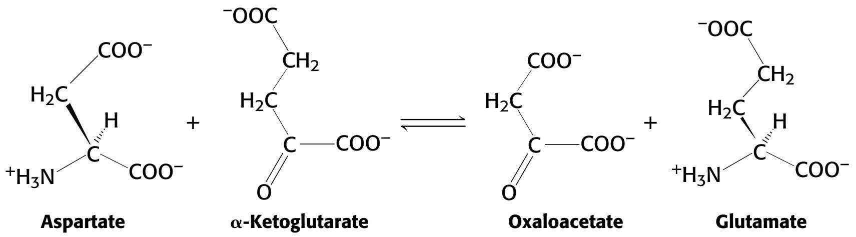

+ ATP
+ ADP
Creatine
Phosphocreatine

Either creatine or ATP may bind first, and either phosphocreatine or ADP may be released first. Phosphocreatine is an important energy source in muscle. Sequential random reactions also can be depicted in the Cleland notation.

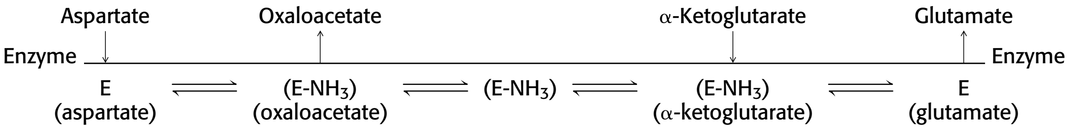

ATP
Creatine
Phosphocreatine
ADP
Enzyme
Enzyme
E (creatine) (ATP)
E (phosphocreatine) (ADP)
Creatine
ATP
ADP
Phosphocreatine

Although the order of certain events is random, the reaction still passes through the ternary complexes including, first, substrates and, then, products.

Double-displacement (ping-pong) reactions. In *double-displacement*, or *ping-pong, reactions,* one or more products are released before all substrates bind the enzyme. The defining feature of double-displacement reactions is the existence of a *substituted enzyme intermediate,* in which the enzyme is temporarily modified. Reactions that shuttle amino groups between amino acids and α-ketoacids are classic examples of double-displacement mechanisms. The enzyme aspartate aminotransferase catalyzes the transfer of an amino group from aspartate to α-ketoglutarate.



The sequence of events can be portrayed as the following Cleland notation:



After aspartate binds to the enzyme, the enzyme accepts aspartate's amino group to form the substituted enzyme intermediate. The first product, oxaloacetate, subsequently departs. The second substrate, α-ketoglutarate, binds to the enzyme, accepts the amino group from the modified enzyme, and is then released as the final product, glutamate. In the Cleland notation, the substrates appear to bounce on and off the enzyme much as a Ping-Pong ball bounces on a table.

Allosteric enzymes do not obey Michaelis–Menten kinetics

The Michaelis–Menten model has greatly assisted the development of enzymology. Its virtues are simplicity and broad applicability. However, the Michaelis–Menten model cannot account for the kinetic properties of many enzymes. An important group of enzymes that do not obey Michaelis–Menten kinetics are the *allosteric enzymes*. These enzymes consist of multiple subunits and multiple active sites.

Allosteric enzymes often display sigmoidal plots of the reaction velocity V_0 versus substrate concentration [S] (Figure 8.13), rather than the hyperbolic plots predicted by the Michaelis–Menten equation (Figure 8.11). In allosteric enzymes, the binding of substrate to one active site can alter the properties of other active sites in the same enzyme molecule. A possible outcome of this interaction between subunits is that the binding of substrate becomes *cooperative;* that is, the binding of substrate to one active site facilitates the binding of substrate to the other active sites. Such cooperativity results in a sigmoidal plot of V_0 versus [S]. In addition, the activity of an allosteric enzyme may be altered by regulatory molecules that reversibly bind to specific sites other than the catalytic sites. The catalytic properties of allosteric enzymes can thus be adjusted to meet the immediate needs of a cell. For this reason, allosteric enzymes are key regulators of metabolic pathways (Chapter 10). Recall that we have already met an allosteric protein, hemoglobin, in Chapter 7.

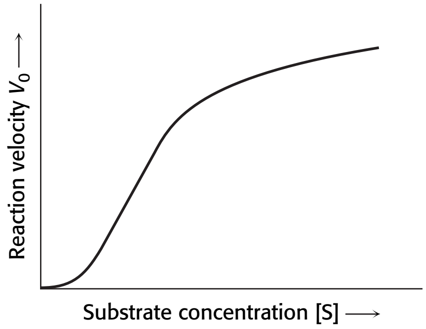


FIGURE 8.13 Kinetics for an allosteric enzyme. Allosteric enzymes display a sigmoidal dependence of reaction velocity on substrate concentration.

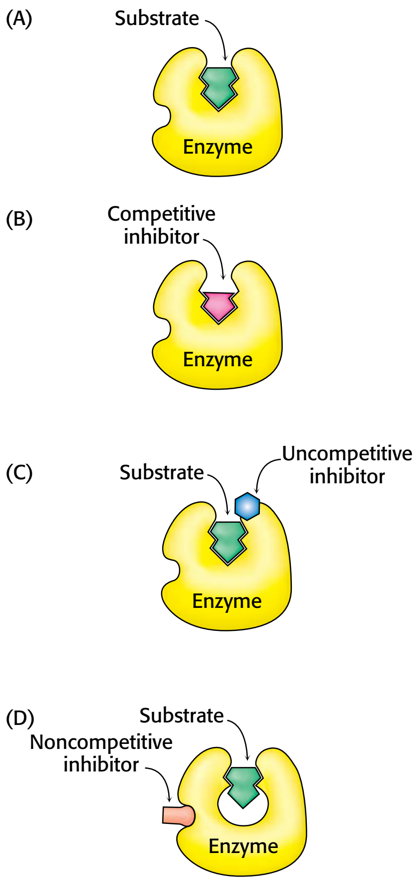


FIGURE 8.14 Distinction between reversible inhibitors. (A) Enzyme–substrate complex; (B) a competitive inhibitor binds at the active site and thus prevents the substrate from binding; (C) an uncompetitive inhibitor binds only to the enzyme–substrate complex; (D) a noncompetitive inhibitor does not prevent the substrate from binding.

8.5 Enzymes Can Be Inhibited by Specific Molecules

The activity of many enzymes can be inhibited by the binding of specific small molecules and ions. This means of inhibiting enzyme activity serves as a major control mechanism in biological systems, typified by the regulation of allosteric enzymes. In addition, many drugs and toxic agents act by inhibiting enzymes (Chapter 36). This type of enzyme inhibition is not usually the result of evolutionary forces, as it is for allosteric enzymes, but rather due to design of inhibitors by scientists or simple chance discovery of inhibitory molecules. Examining inhibition can be a source of insight into the mechanism of enzyme action: specific inhibitors can often be used to identify residues critical for catalysis. Transition-state analogs are especially potent inhibitors.

Enzyme inhibition can be either irreversible or reversible. An *irreversible inhibitor* dissociates very slowly from its target enzyme because it has become tightly bound to the enzyme, either covalently or noncovalently. Some irreversible inhibitors are important drugs. Penicillin acts by covalently modifying the enzyme transpeptidase, thereby preventing the synthesis of bacterial cell walls and thus killing the bacteria (p. 239). Aspirin acts by covalently modifying the enzyme cyclooxygenase, reducing the synthesis of signaling molecules in inflammation.

Reversible inhibition, in contrast with irreversible inhibition, is characterized by a rapid dissociation of the enzyme–inhibitor complex. In the type of reversible inhibition called *competitive inhibition,* an enzyme can bind substrate (forming an ES complex) or inhibitor (EI) but not both (ESI, enzyme–substrate–inhibitor complex). The competitive inhibitor often resembles the substrate and binds to the active site of the enzyme (Figure 8.14). The substrate is thereby prevented from binding to the same active site. *A competitive inhibitor diminishes the rate of catalysis by reducing the proportion of enzyme molecules bound to a substrate.* At any given inhibitor concentration, competitive inhibition can be relieved by increasing the substrate concentration. Under these conditions, the substrate successfully competes with the inhibitor for the active site. Methotrexate is an especially potent competitive inhibitor of the enzyme dihydrofolate reductase, which plays a role in the biosynthesis of purines and pyrimidines. Methotrexate is a structural analog of dihydrofolate, a substrate for dihydrofolate reductase (Figure 8.15). What makes it such a potent competitive inhibitor is that it binds to the enzyme 1000 times as tightly as the natural substrate binds, and it inhibits nucleotide base synthesis. It is used to treat cancer (Section 25.3).

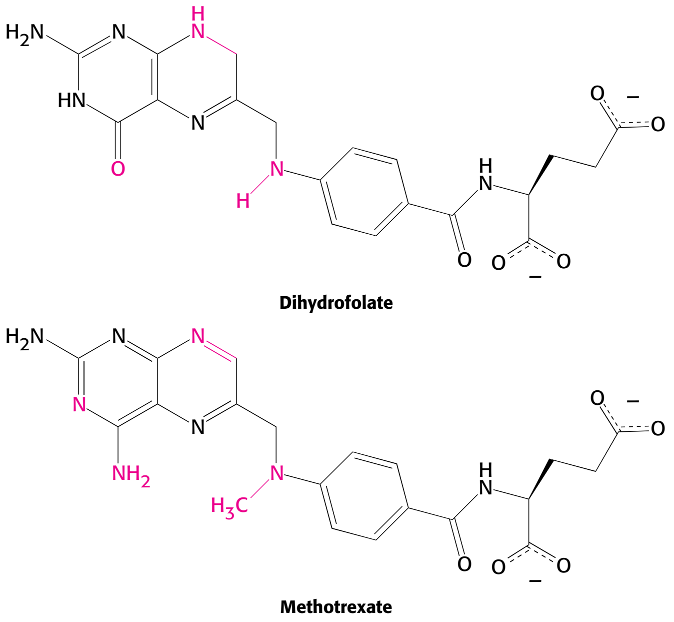


FIGURE 8.15 Enzyme inhibitors. The substrate dihydrofolate and its structural analog methotrexate. Regions with structural differences are shown in red.

Uncompetitive inhibition is essentially substrate-dependent inhibition in that the inhibitor binds only to the enzyme–substrate complex. The binding site of an uncompetitive inhibitor is created only on interaction of the enzyme and substrate (Figure 8.14C). Uncompetitive inhibition cannot be overcome by the addition of more substrate.

In *noncompetitive inhibition,* the inhibitor and substrate can bind simultaneously to an enzyme molecule at different binding sites (Figure 8.14D). Unlike uncompetitive inhibition, a noncompetitive inhibitor can bind free enzyme or the enzyme–substrate complex. A noncompetitive inhibitor acts by decreasing the concentration of functional enzyme rather than by diminishing the proportion of enzyme molecules that are bound to substrate. The net effect is to decrease the turnover number. Noncompetitive inhibition, like uncompetitive inhibition, cannot be overcome by increasing the substrate concentration. A more complex pattern, called *mixed inhibition,* is produced when a single inhibitor both hinders the binding of substrate and decreases the turnover number of the enzyme.

The different types of reversible inhibitors are kinetically distinguishable

How can we determine whether a reversible inhibitor acts by competitive, uncompetitive, or noncompetitive inhibition? Let us consider only enzymes that exhibit Michaelis–Menten kinetics. Measurements of the rates of catalysis at different concentrations of substrate and inhibitor serve to distinguish the three types of inhibition. In *competitive inhibition,* the inhibitor competes with the substrate for the active site. The dissociation constant for the inhibitor is given by

$$K_i = [E][I]/[EI]$$

The smaller the K_i, the more potent the inhibition. The hallmark of competitive inhibition is that it can be overcome by a sufficiently high concentration substrate (Figure 8.16). The effect of a competitive inhibitor is to increase the apparent value of K_M, meaning that more substrate is needed to obtain the same reaction rate. This new value of K_M, called K_M^{app} is numerically equal to

$$K_M^{app} = K_M(1 + [I]/K_i)$$

where [I] is the concentration of inhibitor and K_i is the dissociation constant for the enzyme–inhibitor complex. In the presence of a competitive inhibitor, an enzyme will have the same V_{max} as in the absence of an inhibitor. At a sufficiently high concentration, virtually all the active sites are filled with substrate, and the enzyme is fully operative.

Competitive inhibitors are commonly used as drugs. Drugs such as ibuprofen are competitive inhibitors of enzymes that participate in signaling pathways in the inflammatory response. Statins are drugs that reduce high cholesterol levels by competitively inhibiting a key enzyme in cholesterol biosynthesis (Section 26.3).

In *uncompetitive inhibition,* the inhibitor binds only to the ES complex. This enzyme–substrate–inhibitor complex, ESI, does not go on to form any product. Because some unproductive ESI complex will always be present, V_{max} will be lower in the presence of inhibitor than in its absence (Figure 8.17). The uncompetitive inhibitor lowers the apparent value of K_M because the inhibitor binds to ES to form ESI, depleting ES. To maintain the equilibrium between E and ES, more S binds to E, increasing the apparent value of k_1 and thereby reducing the apparent value of K_M (see equation 18). Thus, a lower concentration of S is required to form half of the maximal concentration of ES. The herbicide glyphosate, also known as Roundup, is an uncompetitive inhibitor of an enzyme in the biosynthetic pathway for aromatic amino acids.

In *noncompetitive inhibition* (Figure 8.18), substrate can bind either

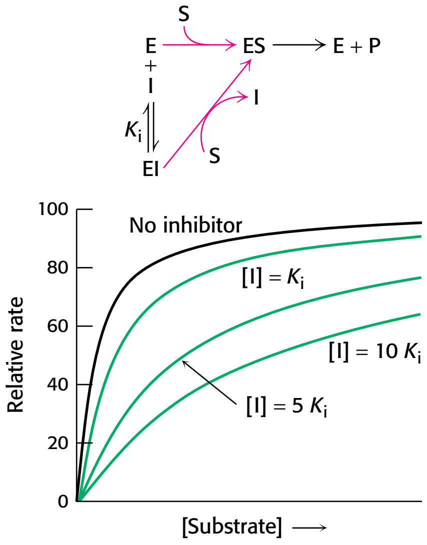


FIGURE 8.16 Kinetics of a competitive inhibitor. As the concentration of a competitive inhibitor increases, higher concentrations of substrate are required to attain a particular reaction velocity. The reaction pathway suggests how sufficiently high concentrations of substrate can completely relieve competitive inhibition.

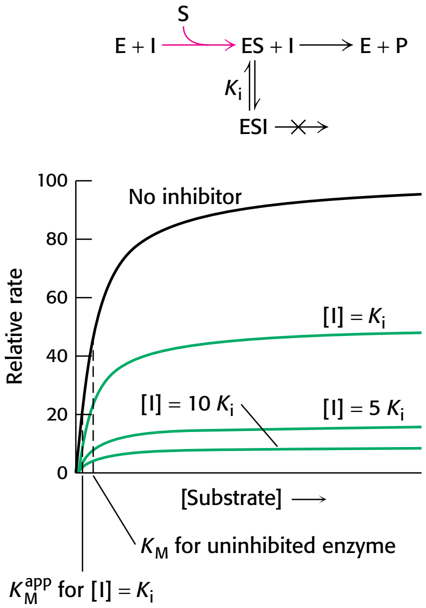


FIGURE 8.17 Kinetics of an uncompetitive inhibitor. The reaction pathway shows that the inhibitor binds only to the enzyme–substrate complex. Consequently, V_{max} cannot be attained, even at high substrate concentrations. The apparent value for K_M is lowered, becoming smaller as more inhibitor is added.

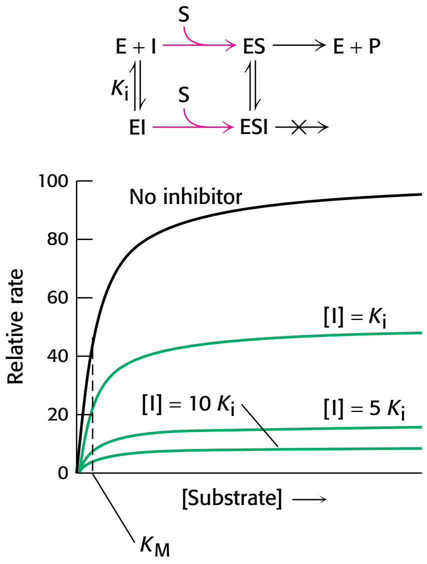


FIGURE 8.18 Kinetics of a noncompetitive inhibitor. The reaction pathway shows that the inhibitor binds both to free enzyme and to an enzyme–substrate complex. Consequently, as with uncompetitive competition, V_{max} cannot be attained. In pure noncompetitive inhibition, K_M remains unchanged, and so the reaction rate increases more slowly at low substrate concentrations than is the case for uncompetitive competition.

to the enzyme or the enzyme–inhibitor complex. However, the enzyme–inhibitor–substrate complex *does not* proceed to form product. In pure noncompetitive inhibition, the K_i for the inhibitor binding to E is the same as for binding to ES complex. The value of V_{max} is decreased to a new value called V_{max}^{app}, whereas the value of K_M is unchanged. The maximal velocity in the presence of a pure noncompetitive inhibitor, V_{max}^{app}, is given by

$$V_{max}^{app} = \frac{V_{max}}{1 + [I]/K_i} \tag{36}$$

Why is V_{max} lowered though K_M remains unchanged? In essence, the inhibitor simply lowers the concentration of functional enzyme. The resulting solution behaves as a more dilute solution of enzyme does. *Noncompetitive inhibition cannot be overcome by increasing the substrate concentration.* Doxycycline, an antibiotic, functions at low concentrations as a noncompetitive inhibitor of a proteolytic enzyme (collagenase). It is used to treat periodontal disease. Some of the toxic effects of lead poisoning may be due to lead's ability to act as a noncompetitive inhibitor of a host of enzymes. Lead reacts with crucial sulfhydryl groups in these enzymes.

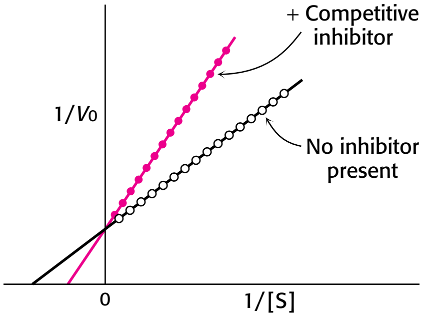


FIGURE 8.19 Competitive inhibition illustrated on a double-reciprocal plot. A double-reciprocal plot of enzyme kinetics in the presence and absence of a competitive inhibitor illustrates that the inhibitor has no effect on V_{max} but increases K_M.

Double-reciprocal plots are especially useful for distinguishing between competitive, uncompetitive, and noncompetitive inhibitors. In competitive inhibition, the intercept on the *y*-axis of the plot of $1/V_0$ versus 1/[S] is the same in the presence and in the absence of inhibitor, although the slope is increased (Figure 8.19). The intercept is unchanged because a competitive inhibitor does not alter V_{max}. The increase in the slope of the $1/V_0$ versus 1/[S] plot indicates the strength of binding of a competitive inhibitor. In the presence of a competitive inhibitor, equation 27 is replaced by

$$\frac{1}{V_0} = \frac{1}{V_{max}} + \frac{K_M}{V_{max}}\left(1 + \frac{[I]}{K_i}\right)\left(\frac{1}{[S]}\right) \tag{37}$$

In other words, the slope of the plot is increased by the factor $(1 + [I]/K_i)$ in the presence of a competitive inhibitor. Consider an enzyme with a K_M of 10^{-4} M. In the absence of inhibitor, when $V_0 = V_{max}/2$ when $[S] = 10^{-4}$ M. In the presence of a 2×10^{-3} M competitive inhibitor that is bound to the enzyme with a K_i of 10^{-3} M, the apparent K_M (K_M^{app}) will be equal to $K_M(1 + [I]/K_i)$, or 3×10^{-4} M. Substitution of these values into equation 37 gives when $V_0 = V_{max}/4$, when $[S] = 10^{-4}$ M. The presence of the competitive inhibitor thus cuts the reaction rate in half at this substrate concentration.

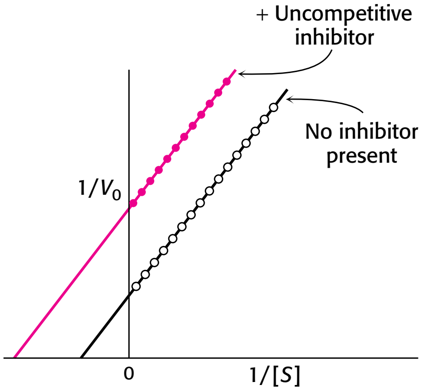


FIGURE 8.20 Uncompetitive inhibition illustrated by a double-reciprocal plot. An uncompetitive inhibitor does not affect the slope of the double-reciprocal plot. V_{max} and K_M are reduced by equivalent amounts.

In uncompetitive inhibition (Figure 8.20), the inhibitor combines only with the enzyme–substrate complex. The equation that describes the double-reciprocal plot for an uncompetitive inhibitor is

$$\frac{1}{V_0} = \frac{K_M}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}}\left(1 + \frac{[I]}{K_i}\right) \tag{38}$$

The slope of the line, $K_M/V_{max,}$ is the same as that for the uninhibited enzyme, but the intercept on the *y*-axis will be increased by $1 + [I]/K_i$. Consequently, the lines in double-reciprocal plots will be parallel.

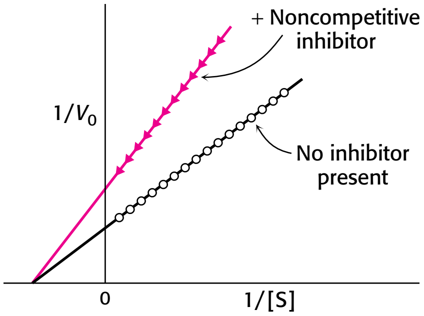


FIGURE 8.21 Noncompetitive inhibition illustrated on a double-reciprocal plot. A double-reciprocal plot of enzyme kinetics in the presence and absence of a pure noncompetitive inhibitor shows that K_M is unaltered and V_{max} is decreased.

In pure noncompetitive inhibition (Figure 8.21), the inhibitor can combine with either the enzyme or the enzyme–substrate complex with the same dissociation constant. The value of V_{max} is decreased to the new value V_{max}^{app}, and so the intercept on the vertical axis is increased (equation 36). The new slope, which is equal to K_M/V_{max}^{app}, is larger by the same factor. In contrast with V_{max}, K_M is not affected by pure noncompetitive inhibition.

Irreversible inhibitors can be used to map the active site

In Chapter 9, we will examine the chemical details of how enzymes function. The first step in obtaining the chemical mechanism of an enzyme is to determine what functional groups are required for enzyme activity. How can we ascertain what these functional groups are? X-ray crystallography of the enzyme bound to its substrate or substrate analog provides one approach. Irreversible inhibitors that covalently bond to the enzyme provide an alternative and often complementary approach: the inhibitors modify the functional groups, which can then be identified. Irreversible inhibitors can be divided into three categories: group-specific reagents, reactive substrate analogs (also called affinity labels), and suicide inhibitors.

Group-specific reagents react with specific side chains of amino acids. An example of a group-specific reagent is diisopropylphosphofluoridate (DIPF). DIPF modifies only 1 of the 28 serine residues in the proteolytic enzyme chymotrypsin and yet inhibits the enzyme, implying that this serine residue is especially reactive. We will see in Chapter 9 that this serine residue is indeed located at the active site. DIPF also revealed a reactive serine residue in acetylcholinesterase, an enzyme important in the transmission of nerve impulses (Figure 8.22). Thus, DIPF and similar compounds that bind and inactivate acetylcholinesterase are potent nerve gases. Most group-specific reagents do not display the exquisite specificity shown by DIPF. Consequently, more specific means of modifying the active site are required.

Affinity labels, or *reactive substrate analogs,* are molecules that are structurally similar to the substrate for an enzyme and that covalently bind to active-site residues. They are thus more specific for the enzyme's active site than are group-specific reagents. Tosyl-L-phenylalanine chloromethyl ketone (TPCK) is a substrate analog for chymotrypsin (Figure 8.23). TPCK binds at the active

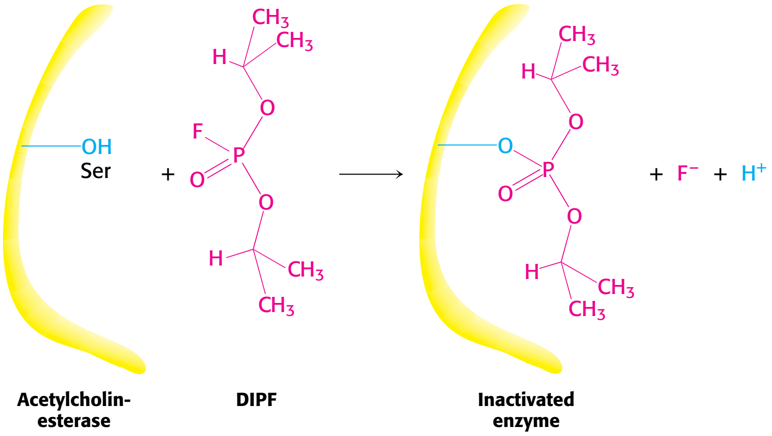

OH
Ser
+
F
P
O
O
O
H
CH₃
CH₃
H
CH₃
CH₃
O
P
O
O
O
+ F⁻ + H⁺
Acetylcholin-esterase
DIPF
Inactivated enzyme

FIGURE 8.22 Enzyme inhibition by diisopropylphosphofluoridate (DIPF), a group-specific reagent. DIPF can inhibit an enzyme by covalently modifying a crucial serine residue.

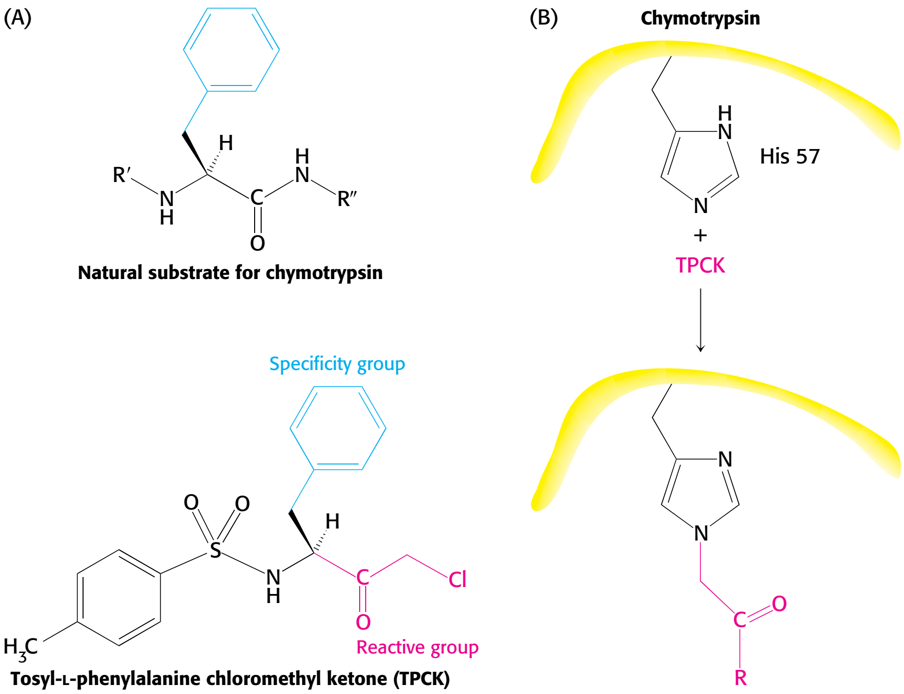

(A)
H
R′
N
H
H
N
R″
C
O
Natural substrate for chymotrypsin
Specificity group
O
O
S
H
N
H
C
Cl
O
H₃C
Reactive group
Tosyl-L-phenylalanine chloromethyl ketone (TPCK)
(B)
Chymotrypsin
H
N
His 57
N
+
TPCK
N
N
C
O
R

FIGURE 8.23 Affinity labeling. (A) Tosyl-L-phenylalanine chloromethyl ketone (TPCK) is a reactive analog of the normal substrate for the enzyme chymotrypsin. (B) TPCK binds at the active site of chymotrypsin and modifies an essential histidine residue.

site and then reacts irreversibly with a histidine residue at that site, inhibiting the enzyme. The compound 3-bromoacetol phosphate is an affinity label for the enzyme triose phosphate isomerase (TPI). It mimics the normal substrate, dihydroxyacetone phosphate, by binding at the active site; then it covalently modifies the enzyme such that the enzyme is irreversibly inhibited (Figure 8.24).

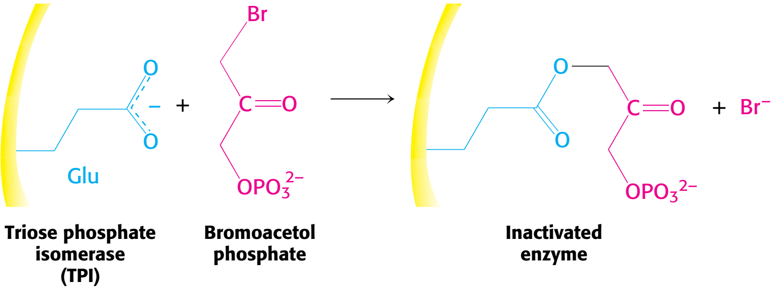
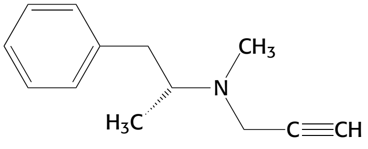


FIGURE 8.24 Bromoacetol phosphate, an affinity label for triose phosphate isomerase (TPI). Bromoacetol phosphate, an analog of dihydroxyacetone phosphate, binds at the active site of the enzyme and covalently modifies a glutamic acid residue required for enzyme activity.

Suicide inhibitors, or *mechanism-based inhibitors,* are modified substrates that provide the most specific means for modifying an enzyme's active site. The inhibitor binds to the enzyme as a substrate and is initially processed by the normal catalytic mechanism. The mechanism of catalysis then generates a chemically reactive intermediate that inactivates the enzyme through covalent modification. The fact that the enzyme participates in its own irreversible inhibition strongly suggests that the covalently modified group on the enzyme is vital for catalysis. An example is *N*,*N*-dimethylpropargylamine, an inhibitor of the enzyme monoamine oxidase (MAO). A flavin prosthetic group of monoamine oxidase oxidizes the *N*,*N*-dimethylpropargylamine, which in turn inactivates the enzyme by binding to N-5 of the flavin prosthetic group (Figure 8.25). Monoamine oxidase deaminates neurotransmitters such as dopamine and serotonin, lowering their levels in the brain. Parkinson disease is associated with low levels of dopamine, and depression is associated with low levels of serotonin. *N*,*N*-Dimethylpropargylamine and (−)deprenyl, another suicide inhibitor of monoamine oxidase, are used to treat Parkinson disease and depression.



(−)Deprenyl

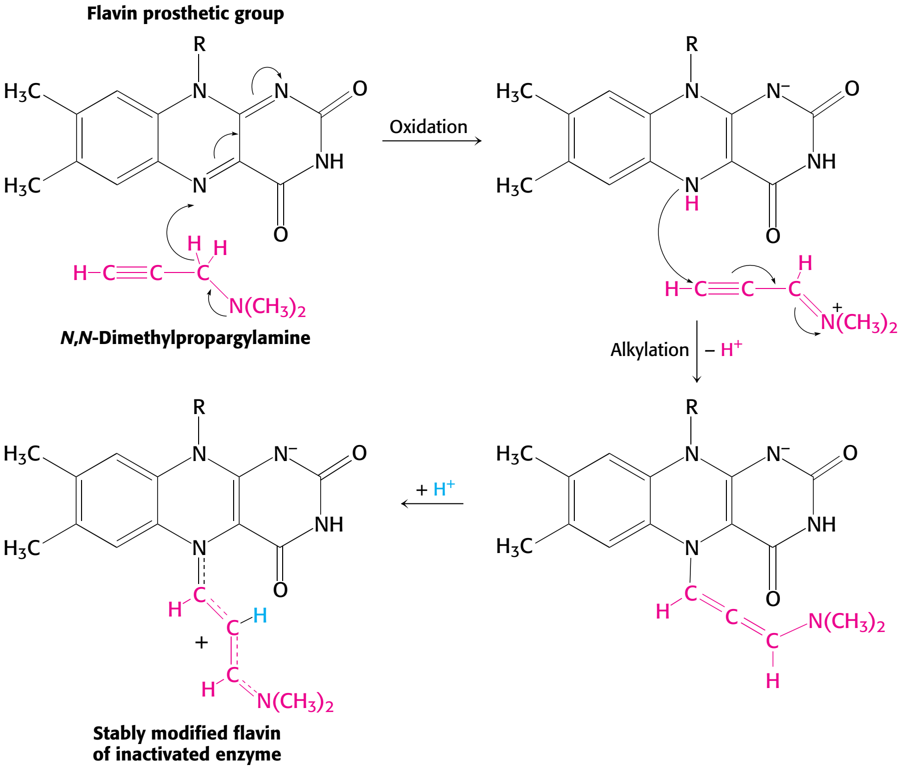


FIGURE 8.25 Mechanism-based (suicide) inhibition. Monoamine oxidase, an enzyme important for neurotransmitter synthesis, requires the cofactor FAD (flavin adenine dinucleotide). *N*,*N*-Dimethylpropargylamine inhibits monoamine oxidase by covalently modifying the flavin prosthetic group only after the inhibitor has been oxidized. The N-5 flavin adduct is stabilized by the addition of a proton. R represents the remainder of the flavin prosthetic group.

Penicillin irreversibly inactivates a key enzyme in bacterial cell-wall synthesis

Penicillin, the first antibiotic discovered, provides us with another example of a clinically useful suicide inhibitor. Penicillin consists of a thiazolidine ring fused to a *β-lactam* ring to which a variable R group is attached by a peptide bond (Figure 8.26A). In benzylpenicillin, for example, R is a benzyl group (Figure 8.26B). This structure can undergo a variety of rearrangements, and, in particular, the β-lactam ring is very labile. Indeed, this instability is closely tied to the antibiotic action of penicillin, as will be evident shortly.

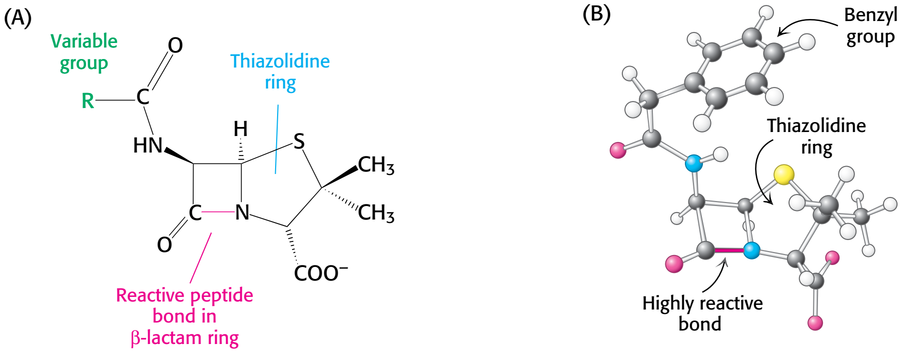

(A)
Variable group
O
R
C
HN
H
Thiazolidine ring
S
CH_3
CH_3
C
N
O
COO^-
Reactive peptide bond in β-lactam ring
(B)
Benzyl group
Thiazolidine ring
Highly reactive bond

FIGURE 8.26 **The reactive site of penicillin is the peptide bond of its β-lactam ring.** (A) Structural formula of penicillin. (B) Representation of benzylpenicillin.

How does penicillin inhibit bacterial growth? Let us consider *Staphylococcus aureus,* the most common cause of staph infections. Penicillin interferes with the synthesis of the *S. aureus* cell wall. The *S. aureus* cell wall is made up of a macromolecule, called a *peptidoglycan* (Figure 8.27), which consists of linear polysaccharide chains that are cross-linked by short peptides (pentaglycines and tetrapeptides). The enormous peptidoglycan molecule confers mechanical support and prevents bacteria from bursting in response to their high internal osmotic pressure. *Glycopeptide transpeptidase* catalyzes the formation of the cross-links that make the peptidoglycan so stable (Figure 8.28). Bacterial cell walls are distinctive in containing D amino acids, which form cross-links by a mechanism different from that used to synthesize proteins.

Penicillin inhibits the cross-linking transpeptidase by the Trojan horse stratagem. The transpeptidase normally forms an *acyl intermediate* with the penultimate D-alanine residue of the D-Ala-D-Ala peptide (Figure 8.29). This covalent acyl-enzyme intermediate then reacts with the amino group of the terminal glycine in another peptide to form the cross-link. Penicillin is welcomed into the active site of the transpeptidase because it mimics the

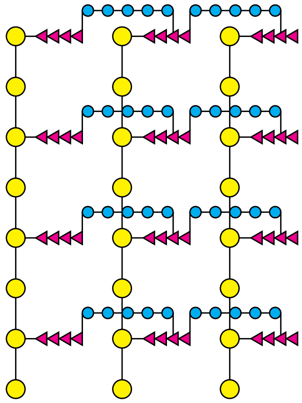


FIGURE 8.27 **Schematic representation of the peptidoglycan in *Staphylococcus aureus.*** The sugars are shown in yellow, the tetrapeptides in red, and the pentaglycine bridges in blue. The cell wall is a single, enormous, bag-shaped macromolecule because of extensive cross-linking.

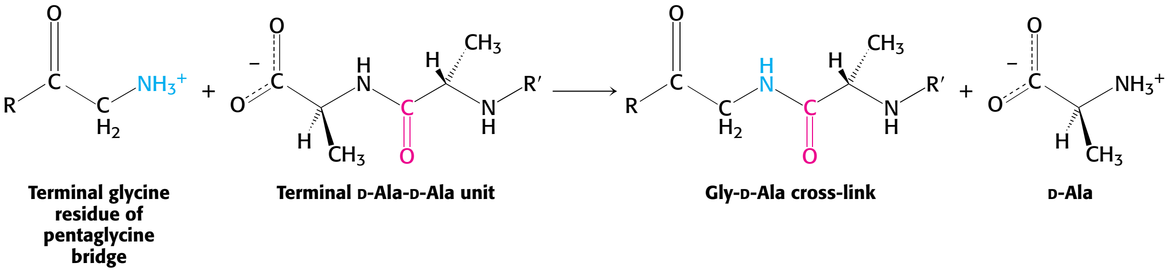
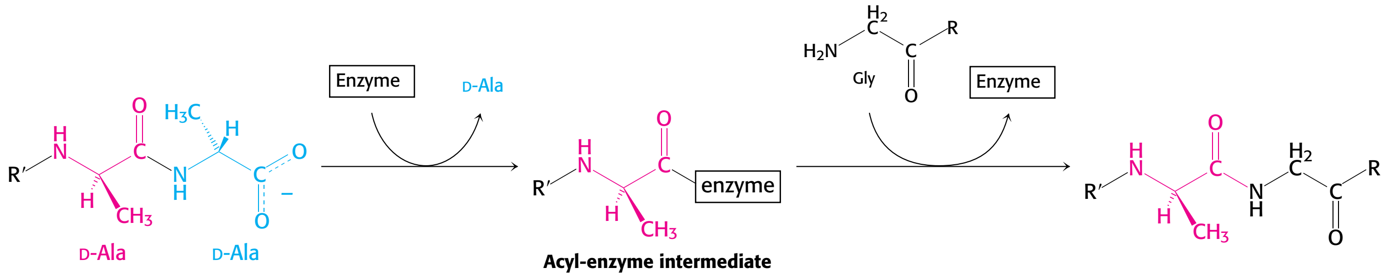

O
R
C
NH_3^+
C
H_2
+
O
−
O
C
H
N
H
CH_3
C
O
N
H
R′
H
CH_3
R
C
O
C
H_2
H
N
C
O
H
CH_3
N
H
R′
+
O
−
O
C
NH_3^+
H
CH_3
Terminal glycine residue of pentaglycine bridge
Terminal D-Ala-D-Ala unit
Gly-D-Ala cross-link
D-Ala

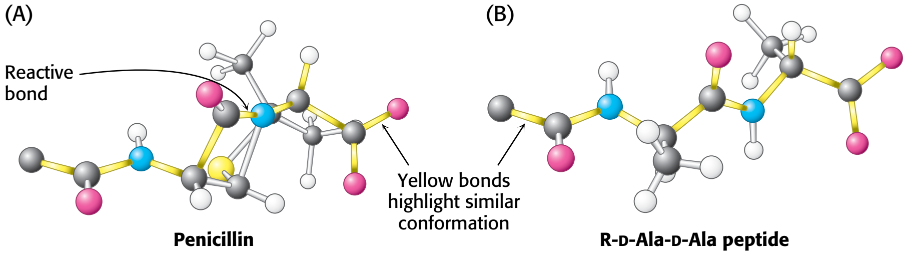
FIGURE 8.28 **Formation of cross-links in *S. aureus* peptidoglycan.** The terminal amino group of the pentaglycine bridge in the cell wall attacks the peptide bond between two D-alanine residues to form a cross-link.



Enzyme
D-Ala
D-Ala
D-Ala
Gly
Enzyme
enzyme
Acyl-enzyme intermediate

FIGURE 8.29 Transpeptidation reaction. An acyl-enzyme intermediate is formed in the transpeptidation reaction leading to cross-link formation.

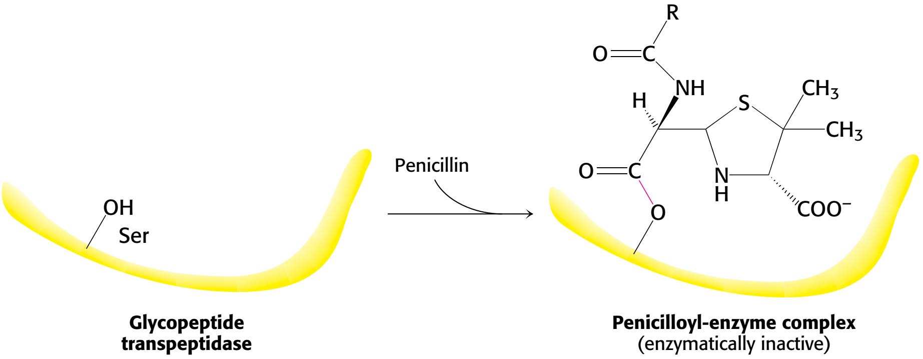
D-Ala-D-Ala moiety of the normal substrate (Figure 8.30). Bound penicillin then forms a covalent bond with a serine residue at the active site of the enzyme. *This penicilloyl-enzyme does not react further. Hence, the transpeptidase is irreversibly inhibited and cell-wall synthesis cannot take place.*



(A)
(B)
Reactive bond
Yellow bonds highlight similar conformation
Penicillin
R-D-Ala-D-Ala peptide

FIGURE 8.30 Conformations of penicillin and a normal substrate. The conformation of penicillin in the vicinity of its reactive peptide bond (A) resembles the postulated conformation of the transition state of R-D-Ala-D-Ala (B) in the transpeptidation reaction. [Information from B. Lee, *J. Mol. Biol.* 61:463–469, 1971.]

Why is penicillin such an effective inhibitor of the transpeptidase? The highly strained, four-membered β-lactam ring of penicillin makes it especially reactive (Figure 8.26). On binding to the transpeptidase, the serine residue at the active site attacks the carbonyl carbon atom of the lactam ring to form the penicilloyl-serine derivative (Figure 8.31). Because the peptidase participates in its own inactivation, penicillin acts as a suicide inhibitor.



OH
Ser
Penicillin
Glycopeptide transpeptidase
Penicilloyl-enzyme complex
(enzymatically inactive)

FIGURE 8.31 Formation of a penicilloyl-enzyme derivative. Penicillin reacts irreversibly with the transpeptidase to inactivate the enzyme.

Transition-state analogs are potent inhibitors of enzymes

We turn now to compounds that provide the most intimate views of the catalytic process itself. Linus Pauling proposed in 1948 that compounds resembling the transition state of a catalyzed reaction should be very effective inhibitors of enzymes. These mimics are called *transition-state analogs*. The inhibition of proline racemase is an instructive example. The racemization

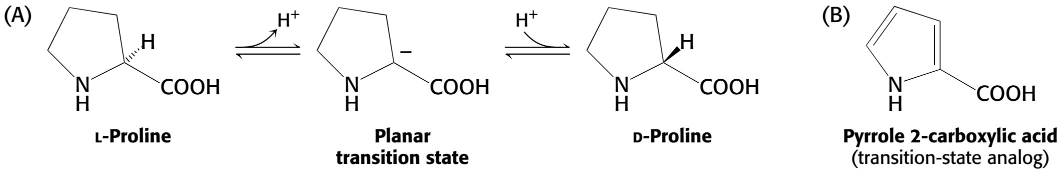


FIGURE 8.32 Inhibition by transition-state analogs. (A) The isomerization of L-proline to D-proline by proline racemase proceeds through a planar transition state in which the α-carbon atom is trigonal rather than tetrahedral. (B) Pyrrole 2-carboxylic acid, a transition-state analog because of its trigonal geometry, is a potent inhibitor of proline racemase.

of proline proceeds through a transition state in which the tetrahedral α-carbon atom has become trigonal (Figure 8.32). In the trigonal form, all three bonds are in the same plane; C_α also carries a net negative charge. This symmetric carbanion can be reprotonated on one side to give the L isomer or on the other side to give the D isomer. This picture is supported by the finding that the inhibitor pyrrole 2-carboxylate binds to the racemase 160 times as tightly as does proline. *The α-carbon atom of this inhibitor, like that of the transition state, is trigonal.* An analog that also carries a negative charge on C_α would be expected to bind even more tightly. In general, highly potent and specific inhibitors of enzymes can be produced by synthesizing compounds that more closely resemble the transition state than the substrate itself. The inhibitory power of transition-state analogs underscores the essence of catalysis: *selective binding of the transition state.*

Catalytic antibodies demonstrate the importance of selective binding of the transition state to enzymatic activity

Recall that antibodies bind precisely to their targets (antigens), and that specific antibodies can be generated against any antigen (Section 3.3). Antibodies that recognize transition states should function as catalysts, if our understanding of the importance of the transition state to catalysis is correct. The preparation of an antibody that catalyzes the insertion of a metal ion into a porphyrin nicely illustrates the validity of this approach. Ferrochelatase, the final enzyme in the biosynthetic pathway for the production of heme, catalyzes the insertion of Fe^{2+} into protoporphyrin IX (Section 24.4). The nearly planar porphyrin must be bent for iron to enter.

The challenge was to find a transition-state analog for this metallation reaction that could be used as an antigen (immunogen) to generate an antibody. The solution came from studies showing that an alkylated porphyrin, *N*-methylmesoporphyrin, is a potent inhibitor of ferrochelatase (Figure 8.33). This compound resembles the transition state because N-*alkylation forces the porphyrin to be bent*. Moreover, *N*-alkylporphyrins were known to chelate metal ions 10^4 times as fast as their unalkylated counterparts do. Bending increases the exposure of the pyrrole nitrogen lone pairs of electrons to solvent, which enables the binding of the iron ion.

An antibody catalyst was produced with the use of an *N*-alkylporphyrin as the antigen. The resulting antibody presumably distorts a planar porphyrin to facilitate the entry of a metal ion. On average, an antibody molecule metallated 80 porphyrin molecules per hour, a rate only 10-fold less than that of ferrochelatase, and 2500-fold faster than the uncatalyzed reaction. *Catalytic antibodies (abzymes) can indeed be produced by using transition-state analogs as antigens.* Antibodies catalyzing many other kinds of chemical reactions such as ester and amide hydrolysis have been produced by using similar strategies. Studies with transition-state analogs provide strong

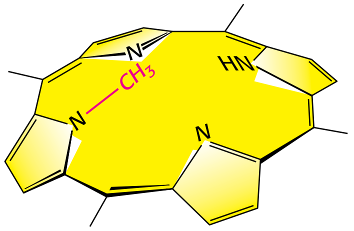


FIGURE 8.33 *N*-Methylmesoporphyrin is a transition-state analog used to generate catalytic antibodies. The insertion of a metal ion into a porphyrin by ferrochelatase proceeds through a transition state in which the porphyrin is bent. *N*-Methylmesoporphyrin, a bent porphyrin that resembles the transition state of the ferrochelatase-catalyzed reaction, was used to generate an antibody that also catalyzes the insertion of a metal ion into a porphyrin ring.

evidence that enzymes function by assuming a conformation in the active site that is complementary in structure to the transition state. *The power of transition-state analogs is now evident: (1) they are sources of insight into catalytic mechanisms, (2) they can serve as potent and specific inhibitors of enzymes, and (3) they can be used as immunogens to generate a wide range of novel catalysts.*

8.6 Enzymes Can Be Studied One Molecule at a Time

Most experiments that are performed to determine an enzyme characteristic require an enzyme preparation in a buffered solution. Even a few microliters of such a solution will contain millions of enzyme molecules. Much that we have learned about enzymes thus far has come from such experiments, called *ensemble studies*. A basic assumption of ensemble studies is that all of the enzyme molecules are the same or very similar. When we determine an enzymatic property such as the value of K_M in ensemble studies, that value is of necessity an average value of all of the enzyme molecules present. However, as discussed in Chapter 2, we now know that molecular heterogeneity, the ability of a molecule, over time, to assume several different structures that differ slightly in stability, is an inherent property of all large biomolecules (p. 53). How can we tell if this molecular heterogeneity affects enzyme activity?

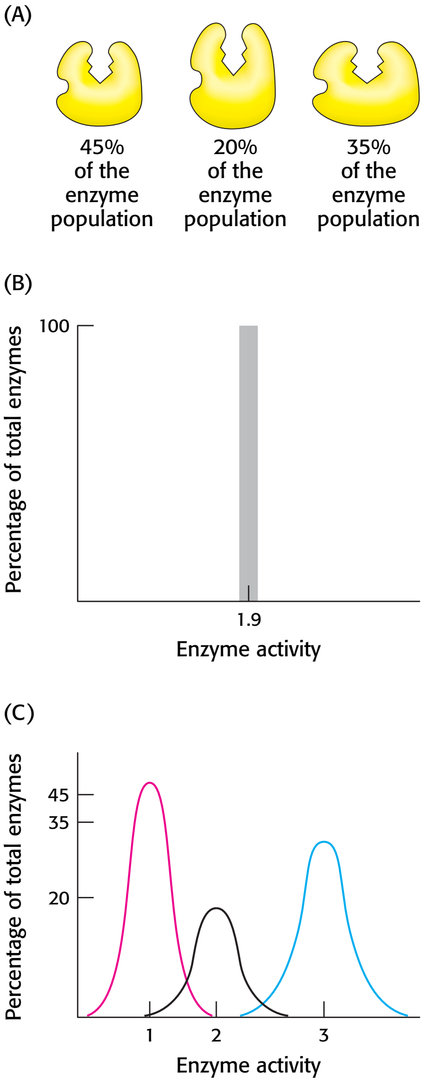


FIGURE 8.34 Single molecule studies can reveal molecular heterogeneity. (A) Complex biomolecules, such as enzymes, display molecular heterogeneity. (B) When measuring an enzyme property using ensemble methods, an average value of all of the enzymes present is the result. (C) Single enzyme studies reveal molecular heterogeneity, with the various forms showing different properties.

By way of example, consider a hypothetical situation. A Martian visits Earth to learn about higher education. The spacecraft hovers high above a university, and our Martian meticulously records how the student population moves about campus. Much information can be gathered from such studies: where students are likely to be at certain times on certain days, which buildings are used when and by how many. Now, suppose our visitor developed a high-magnification camera that could follow one student throughout the day. Such data would provide a much different perspective on college life: What does this student eat? To whom does she talk? How much time does she spend studying? This new *in singulo* method, examining one individual at a time, yields a great deal of new information but also illustrates a potential pitfall of studying individuals, be they students or enzymes: How can we be certain that the student or molecule is representative and not an outlier? This pitfall can be overcome by studying enough individuals to satisfy statistical analysis for validity.

Let us leave our Martian to his observations, and consider a more biochemical situation. Figure 8.34A shows an enzyme that displays molecular heterogeneity, with three active forms that catalyze the same reaction but at different rates. These forms have slightly different stabilities, but thermal noise is sufficient to interconvert the forms. Each form is present as a fraction of the total enzyme population as indicated. If we were to perform an experiment to determine enzyme activity under a particular set of conditions with the use of ensemble methods, we would get a single value, which would represent the average of the heterogeneous assembly (Figure 8.34B). However, were we to perform a sufficient number of single-molecule experiments, we would discover that the enzyme has three different molecular forms with very different activities (Figure 8.34C). Moreover, these different forms would most likely correspond to important biochemical differences.

The development of powerful techniques—such as patch-clamp recording, single-molecule fluorescence, and optical tweezers—has enabled biochemists to look into the workings of individual molecules.

We will examine single-molecule studies of membrane channels with the use of patch-clamp recording (Section 13.4), ATP-synthesizing complexes with the use of single-molecule fluorescence (Section 18.4) and molecular motors with the use of an optical trap (Section 34.2). We are now able to observe events at a molecular level that reveal rare or transient structures and fleeting events in a reaction sequence, as well as to measure mechanical forces affecting or generated by an enzyme. Single-molecule studies open a new vista on the function of enzymes in particular and on all large biomolecules in general.

SUMMARY

8.1 Enzymes Are Powerful and Highly Specific Catalysts

Most catalysts in biological systems are enzymes, and nearly all enzymes are proteins. Enzymes are highly specific and have great catalytic power. They can enhance reaction rates by factors of 10^6 or more. Many enzymes require cofactors for activity. Such cofactors can be metal ions or small, vitamin-derived organic molecules called coenzymes.

8.2 Gibbs Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes

Free energy (G) is the most valuable thermodynamic function for understanding the energetics of catalysis. A reaction can take place spontaneously only if the change in free energy (ΔG) is negative. The free-energy change of a reaction that takes place when reactants and products are at unit activity is called the standard free-energy change (ΔG°). Biochemists use $\Delta G^{\circ\prime}$, the standard free-energy change at pH 7. Enzymes do not alter reaction equilibria; rather, they increase the rate at which equilibrium is attained.

8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

Enzymes serve as catalysts by decreasing the free energy of activation of chemical reactions. Enzymes accelerate reactions by providing a reaction pathway in which the transition state (the highest-energy species) has a lower free energy and hence is more rapidly formed than in the uncatalyzed reaction.

The first step in catalysis is the formation of an enzyme–substrate complex. Substrates are bound to enzymes at active-site clefts from which water is largely excluded when the substrate is bound. The specificity of enzyme–substrate interactions arises mainly from hydrogen bonding, which is directional, and from the shape of the active site, which rejects molecules that do not have a sufficiently complementary shape. Enzymes facilitate formation of the transition state by a dynamic process in which the substrate binds to specific conformations of the enzyme, accompanied by conformational changes at active sites that result in catalysis.

8.4 The Michaelis–Menten Model Accounts for the Kinetic Properties of Many Enzymes

The kinetic properties of many enzymes are described by the Michaelis–Menten model. In this model, an enzyme (E) combines with a substrate (S) to form an enzyme–substrate (ES) complex, which can proceed to form a product (P) or to dissociate into E and S.

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

The rate of formation of product V_0 is given by the Michaelis–Menten equation:

$$V_0 = V_{\max}\frac{[\mathrm{S}]}{[\mathrm{S}] + K_{\mathrm{M}}}$$

in which $V_{\max}$ is the reaction rate when the enzyme is fully saturated with substrate and K_{M}, the Michaelis constant, is the substrate concentration at which the reaction rate is half maximal. The maximal rate, $V_{\max}$, is equal to the product of k_2, or k_{cat}, and the total concentration of enzyme. The kinetic constant k_{cat} called the turnover number, is the number of substrate molecules converted into product per unit time at a single catalytic site when the enzyme is fully saturated with substrate. Turnover numbers for most enzymes are between 1 and 10^4 per second. The ratio of $k_{\mathrm{cat}}/K_{\mathrm{M}}$ provides a measure of enzyme efficiency and specificity.

Allosteric enzymes constitute an important class of enzymes whose catalytic activity can be regulated. These enzymes, which do not conform to Michaelis–Menten kinetics, have multiple active sites. These active sites display cooperativity, as evidenced by a sigmoidal dependence of reaction velocity on substrate concentration.

8.5 Enzymes Can Be Inhibited by Specific Molecules

Specific small molecules or ions can inhibit even nonallosteric enzymes. In irreversible inhibition, the inhibitor is covalently linked to the enzyme or bound so tightly that its dissociation from the enzyme is very slow. Covalent inhibitors provide a means of mapping the enzyme's active site. In contrast, reversible inhibition is characterized by a more rapid and less stable interaction between enzyme and inhibitor. A competitive inhibitor prevents the substrate from binding to the active site. It reduces the reaction velocity by diminishing the proportion of enzyme molecules that are bound to substrate. Competitive inhibition can be overcome by raising the substrate concentration. In uncompetitive inhibition, the inhibitor combines only with the enzyme–substrate complex. In noncompetitive inhibition, the inhibitor decreases the turnover number. Uncompetitive and noncompetitive inhibition cannot be overcome by raising the substrate concentration.

The essence of catalysis is selective stabilization of the transition state. Hence, an enzyme binds the transition state more tightly than it binds the substrate. Transition-state analogs are stable compounds that mimic key features of this highest-energy species. They are potent and specific inhibitors of enzymes. Proof that transition-state stabilization is a key aspect of enzyme activity comes from the generation of catalytic antibodies. Transition-state analogs are used as antigens, or immunogens, in generating catalytic antibodies.

8.6 Enzymes Can Be Studied One Molecule at a Time

Many enzymes are now being studied *in singulo,* at the level of a single molecule. Such studies are important because they yield information that is difficult to obtain in studies of populations of molecules. Single-molecule methods reveal a distribution of enzyme characteristics rather than an average value as is acquired with the use of ensemble methods.

APPENDIX: Enzymes are Classified on the Basis of the Types of Reactions That They Catalyze

Many enzymes have common names that provide little information about the reactions that they catalyze. For example, a proteolytic enzyme secreted by the pancreas is called trypsin. Most other enzymes are named for their substrates and for the reactions that they catalyze, with the suffix "ase" added. Thus, a peptide hydrolase is an enzyme that hydrolyzes peptide bonds, whereas ATP synthase is an enzyme that synthesizes ATP.

To bring some consistency to the classification of enzymes, in 1964 the International Union of Biochemistry established an Enzyme Commission to develop a nomenclature for enzymes. Reactions were divided into six major groups numbered 1 through 6 (Table 8.8). These groups were subdivided and further subdivided so that a four-number code preceded by the letters *EC* for Enzyme Commission could precisely identify all enzymes.

Consider as an example nucleoside monophosphate (NMP) kinase, an enzyme that we will examine in detail in Section 9.4. It catalyzes the following reaction:

$$\mathrm{ATP} + \mathrm{NMP} \rightleftharpoons \mathrm{ADP} + \mathrm{NDP}$$

NMP kinase transfers a phosphoryl group from ATP to any nucleotide monophosphate (NMP) to form a nucleoside diphosphate (NDP) and ADP. Consequently, it is a transferase, or member of group 2. Many groups other than phosphoryl groups, such as sugars and single-carbon units, can be transferred. Transferases that shift a phosphoryl group are designated 2.7. Various functional groups can accept the phosphoryl group. If a phosphate is the acceptor, the transferase is designated 2.7.4. The final number designates the acceptor more precisely. In regard to NMP kinase, a nucleoside monophosphate is the acceptor, and the enzyme's designation is EC 2.7.4.4. Although the common names are used routinely, the classification number is used when the precise identity of the enzyme is not clear from the common name alone.

TABLE 8.8 Six major classes of enzymes

Class	Type of reaction	Example	Chapter
1. Oxidoreductases	Oxidation–reduction	Lactate dehydrogenase	16
2. Transferases	Group transfer	Nucleoside monophosphate kinase (NMP kinase)	9
3. Hydrolases	Hydrolysis reactions (transfer of functional groups to water)	Chymotrypsin	9
4. Lyases	Addition or removal of groups to form double bonds	Fumarase	17
5. Isomerases	Isomerization (intramolecular group transfer)	Triose phosphate isomerase	16
6. Ligases	Ligation of two substrates at the expense of ATP hydrolysis	Aminoacyl-tRNA synthetase	30

KEY TERMS

enzyme (p. 216)
substrate (p. 216)
cofactor (p. 217)
apoenzyme (p. 217)
holoenzyme (p. 217)
coenzyme (p. 217)
prosthetic group (p. 217)
free energy (p. 218)
transition state (p. 221)
free energy of activation (p. 221)
active site (p. 222)
induced fit (p. 224)
K_M (the Michaelis constant) (p. 227)
V_{max} (maximal rate) (p. 228)
Michaelis–Menten equation (p. 228)
Lineweaver–Burk equation (double-reciprocal plot) (p. 229)
turnover number (p. 230)
k_{cat}/K_M ratio (the specificity constant) (p. 230)
sequential reaction (p. 232)
double-displacement (ping-pong) reaction (p. 233)
allosteric enzyme (p. 233)
competitive inhibition (p. 234)
uncompetitive inhibition (p. 234)
noncompetitive inhibition (p. 234)
group-specific reagent (p. 237)
affinity label (reactive substrate analog) (p. 237)
mechanism-based (suicide) inhibition (p. 238)
transition-state analog (p. 240)
catalytic antibody (abzyme) (p. 241)

PROBLEMS

1. *Raisons d'être*. What are the two properties of enzymes that make them especially useful catalysts?

2. *Partners*. What does an apoenzyme require to become a holoenzyme?

3. *Different partners*. What are the two main types of cofactors?

4. *One a day*. Why are vitamins necessary for good health?

5. *A function of state*. What is the fundamental mechanism by which enzymes enhance the rate of chemical reactions?

6. *Nooks and crannies*. What is the structural basis for enzyme specificity?

7. *Give with one hand, take with the other*. Why does the activation energy of a reaction not appear in the final ΔG of the reaction?

8. *The more things change, the more they stay the same*. Suppose that, in the absence of enzyme, the forward rate constant (k_F) for the conversion of S into P is $10^{-4}\ \text{s}^{-1}$ and the reverse rate constant (k_R) for the conversion of P into S is $10^{-6}\ \text{s}^{-1}$.

$$\text{S} \underset{10^{-6}\,\text{s}^{-1}}{\overset{10^{-4}\,\text{s}^{-1}}{\rightleftharpoons}} \text{P}$$

(a) What is the equilibrium for the reaction? What is the $\Delta G°'$?

(b) Suppose an enzyme enhances the rate of the reaction 100 fold. What are the rate constants for the enzyme-catalyzed reaction? The equilibrium constant? The $\Delta G°'$?

9. *Mountain climbing*. Proteins are thermodynamically unstable. The ΔG of the hydrolysis of proteins is quite negative, yet proteins can be quite stable. Explain this apparent paradox. What does it tell you about protein synthesis?

10. *Protection*. Suggest why the enzyme lysozyme, which degrades cell walls of some bacteria, is present in tears.

11. *Mutual attraction*. What is meant by the term *binding energy?*

12. *Catalytically binding*. What is the role of binding energy in enzyme catalysis?

13. *Sticky situation*. What would be the result of an enzyme having a greater binding energy for the substrate than for the transition state?

14. *Stability matters*. Transition-state analogs, which can be used as enzyme inhibitors and to generate catalytic antibodies, are often difficult to synthesize. Suggest a reason.

15. *Match'em*. Match the K'_{eq} values with the appropriate $\Delta G°'$ values.

	K'_{eq}	$\Delta G°'(\text{kJ mol}^{-1})$
(a)	1	28.53
(b)	10^{-5}	−11.42
(c)	10^{4}	5.69
(d)	10^{2}	0
(e)	10^{-1}	−22.84

16. *Free energy!* Assume that you have a solution of 0.1 M glucose 6-phosphate. To this solution, you add the enzyme phosphoglucomutase, which catalyzes the following reaction:

$$\text{Glucose 6-phosphate} \xrightleftharpoons{\text{Phosphoglucomutase}} \text{glucose 1-phosphate}$$

The $\Delta G°'$ for the reaction is $+7.5\ \text{kJ mol}^{-1}$ ($+1.8\ \text{kcal mol}^{-1}$).

(a) Does the reaction proceed as written? If so, what are the final concentrations of glucose 6-phosphate and glucose 1-phosphate?

(b) Under what cellular conditions could you produce glucose 1-phosphate at a high rate?

17. *Free energy, too!* Consider the following reaction:

$$\text{Glucose 6-phosphate} \xrightleftharpoons{\text{Phosphoglucomutase}} \text{glucose 1-phosphate}$$

After reactant and product were mixed and allowed to reach equilibrium at 25°C, the concentration of each compound was measured:

$$[\text{Glucose 1-phosphate}]_{eq} = 0.01\ \text{M}$$
$$[\text{Glucose 6-phosphate}]_{eq} = 0.19\ \text{M}$$

Calculate K_{eq} and $\Delta G°'$.

18. *Keeping busy*. Many isolated enzymes, if incubated at 37°C, will be denatured. However, if the enzymes are incubated at 37°C in the presence of substrate, the enzymes are catalytically active. Explain this apparent paradox.

19. *Active yet responsive*. What is the biochemical advantage of having a K_M approximately equal to the substrate concentration normally available to an enzyme?

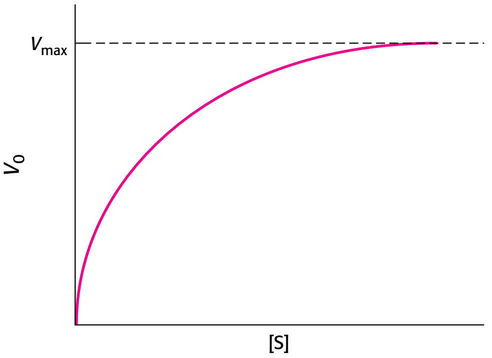
20. *Affinity or not affinity? That is the question*. The affinity between a protein and a molecule that binds to the protein is frequently expressed in terms of a dissociation constant K_d.

$$\text{Protein–small molecule complex} \rightleftharpoons \text{protein} + \text{small molecule}$$

$$K_d = \frac{[\text{protein}][\text{small molecule}]}{[\text{Protein} - \text{small molecule complex}]}$$

Does K_M measure the affinity of the enzyme complex? Under what circumstances might K_M approximately equal K_d?

21. *Angry biochemists.* Many biochemists go bananas, and justifiably, when they see a Michaelis–Menten plot like the one shown below. To see why, determine the V_0 as a fraction of V_{max} when the substrate concentration is equal to 10 K_M and 20 K_M. Please control your outrage.



22. *Hydrolytic driving force.* The hydrolysis of pyrophosphate to orthophosphate is important in driving forward biosynthetic reactions such as the synthesis of DNA. This hydrolytic reaction is catalyzed in *E. coli* by a pyrophosphatase that has a mass of 120 kDa and consists of six identical subunits. For this enzyme, a unit of activity is defined as the amount of enzyme that hydrolyzes 10 μmol of pyrophosphate in 15 minutes at 37°C under standard assay conditions. The purified enzyme has a V_{max} of 2800 units per milligram of enzyme.

(a) How many moles of substrate is hydrolyzed per second per milligram of enzyme when the substrate concentration is much greater than K_M?

(b) How many moles of active sites are there in 1 mg of enzyme? Assume that each subunit has one active site.

(c) What is the turnover number of the enzyme? Compare this value with others mentioned in this chapter.

23. *Destroying the Trojan horse.* Penicillin is hydrolyzed and thereby rendered inactive by penicillinase (also known as β-lactamase), an enzyme present in some penicillin-resistant bacteria. The mass of this enzyme in *Staphylococcus aureus* is 29.6 kDa. The amount of penicillin hydrolyzed in 1 minute in a 10-ml solution containing 10^{-9} g of purified penicillinase was measured as a function of the concentration of penicillin. Assume that the concentration of penicillin does not change appreciably during the assay.

[Penicillin] μM	Amount hydrolyzed (nmol)
1	0.11
3	0.25
5	0.34
10	0.45
30	0.58
50	0.61

(a) Plot V_0 versus [S] and $1/V_0$ versus 1/[S] for these data. Does penicillinase appear to obey Michaelis–Menten kinetics? If so, what is the value of K_M?

(b) What is the value of V_{max}?

(c) What is the turnover number of penicillinase under these experimental conditions? Assume one active site per enzyme molecule.

24. *Counterpoint.* Penicillinase (β-lactamase) hydrolyzes penicillin. Compare penicillinase with glycopeptide transpeptidase.

25. *A different mode.* The kinetics of an enzyme is measured as a function of substrate concentration in the presence and absence of 100 μM inhibitor.

(a) What are the values of V_{max} and K_M in the presence of this inhibitor?

(b) What type of inhibition is it?

(c) What is the dissociation constant of this inhibitor?

[S] (μM)	Velocity (μmol minute^{-1})	
	No inhibitor	Inhibitor
3	10.4	2.1
5	14.5	2.9
10	22.5	4.5
30	33.8	6.8
90	40.5	8.1

(d) If [S] = 30 μM, what fraction of the enzyme molecules have a bound substrate in the presence and in the absence of 100 μM inhibitor?

26. *A fresh view.* The plot of $1/V_0$ versus 1/[S] is sometimes called a Lineweaver–Burk plot. Another way of expressing the kinetic data is to plot V_0 versus V_0/[S], which is known as an Eadie–Hofstee plot.

(a) Rearrange the Michaelis–Menten equation to give V_0 as a function of V_0/[S].

(b) What is the significance of the slope, the *y*-intercept, and the *x*-intercept in a plot of V_0 versus V_0/[S]?

(c) Sketch a plot of V_0 versus V_0/[S] in the absence of an inhibitor, in the presence of a competitive inhibitor, and in the presence of a noncompetitive inhibitor.

27. *Defining attributes.* What is the defining characteristic for an enzyme catalyzing a sequential reaction? A double-displacement reaction?

28. *Competing substrates.* Suppose that two substrates, A and B, compete for an enzyme. Derive an expression relating the ratio of the rates of utilization of A and B, V_A/V_B, to the concentrations of these substrates and their values of k_{cat} and K_M. (Hint: Express V_A as a function of k_{cat}/K_M for

substrate A, and do the same for V_B.) Is specificity determined by K_M alone?

29. *A tenacious mutant.* Suppose that a mutant enzyme binds a substrate 100 times as tightly as does the native enzyme. What is the effect of this mutation on catalytic rate if the binding of the transition state is unaffected?

30. *More Michaelis–Menten.* For an enzyme that follows simple Michaelis–Menten kinetics, what is the value of $V_{\max}$ if V_0 is equal to 1 μmol minute^{-1} at 10 K_M?

31. *Controlled paralysis.* Succinylcholine is a fast-acting, short-duration muscle relaxant that is used when a tube is inserted into a patient's trachea or when a bronchoscope is used to examine the trachea and bronchi for signs of cancer. Within seconds of the administration of succinylcholine, the patient experiences muscle paralysis and is placed on a respirator while the examination proceeds. Succinylcholine is a competitive inhibitor of acetylcholinesterase, a nervous system enzyme, and this inhibition causes paralysis. However, succinylcholine is hydrolyzed by blood-serum cholinesterase, which shows a broader substrate specificity than does the nervous system enzyme. Paralysis lasts until the succinylcholine is hydrolyzed by the serum cholinesterase, usually several minutes later.

(a) As a safety measure, serum cholinesterase is measured before the examination takes place. Explain why this measurement is good idea.

(b) What would happen to the patient if the serum cholinesterase activity were only 10 units of activity per liter rather than the normal activity of about 80 units?

(c) Some patients have a mutant form of the serum cholinesterase that displays a K_M of 10 mM, rather than the normal 1.4 mM. What will be the effect of this mutation on the patient?

Data Interpretation Problems

32. *A natural attraction, but more complicated.* You have isolated two versions of the same enzyme, a wild type and a mutant differing from the wild type at a single amino acid. Working carefully but expeditiously, you then establish the following kinetic characteristics of the enzymes.

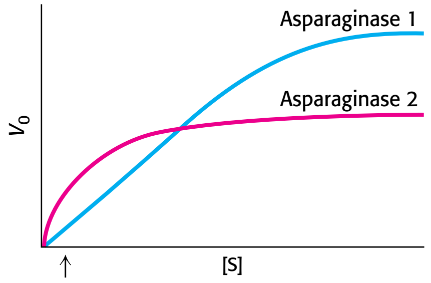
	Maximum velocity	K_M
Wild type	100 μmol/min	10 mM
Mutant	1 μmol/min	0.1 mM

(a) With the assumption that the reaction occurs in two steps in which k_{-1} is much larger than k_2, which enzyme has the higher affinity for substrate?

(b) What is the initial velocity of the reaction catalyzed by the wild-type enzyme when the substrate concentration is 10 mM?

(c) Which enzyme alters the equilibrium more in the direction of product?

33. *K_M matters.* The amino acid asparagine is required by cancer cells to proliferate. Treating patients with the enzyme asparaginase is sometimes used as a chemotherapy treatment. Asparaginase hydrolyzes asparagine to aspartate and ammonia. The adjoining illustration shows the Michaelis–Menten curves for two asparaginases from different sources, as well as the concentration of asparagine in the environment (indicated by the arrow). Which enzyme would make a better chemotherapeutic agent?



34. *Enzyme specificity.* Catalysis of the cleavage of peptide bonds in small peptides by a proteolytic enzyme is described in the following table.

Substrate	K_M (mM)	k_{cat} (s^{-1})
EMTA↓G	4.0	24
EMTA↓A	1.5	30
EMTA↓F	0.5	18

The arrow indicates the peptide bond cleaved in each case.

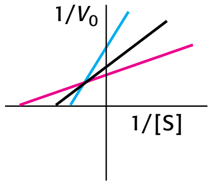
(a) If a mixture of these peptides were presented to the enzyme with the concentration of each peptide being the same, which peptide would be digested most rapidly? Most slowly? Briefly explain your reasoning, if any.

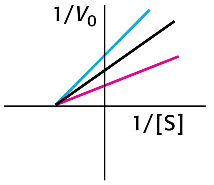
(b) The experiment is performed again on another peptide with the following results.

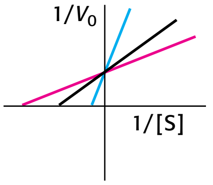
EMTI↓F	9	18

On the basis of these data, suggest the features of the amino acid sequence that dictate the specificity of the enzyme.

35. *Varying the enzyme.* For a one-substrate, enzyme-catalyzed reaction, double-reciprocal plots were determined for three different enzyme concentrations. Which of the following three families of curve would you expect to be obtained? Explain.



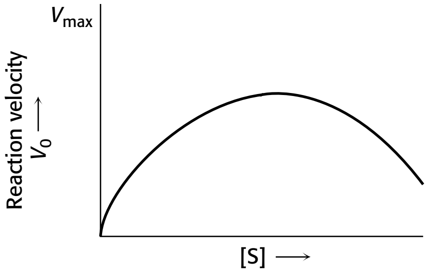




36. *Mental experiment.* Picture in your mind the velocity vs. substrate concentration curve for a typical Michaelis-Menten enzyme. Now, imagine that the experimental conditions are altered as described below. For each of the conditions described, fill in the table indicating precisely (when possible) the effect on $V_{\max}$ and K_M of the imagined Michaelis-Menten enzyme.

Experimental condition	$V_{\max}$	K_M
a. Twice as much enzyme is used.		
b. Half as much enzyme is used		
c. A competitive inhibitor is present.		
d. An uncompetitive inhibitor is present.		
e. A pure non-competitive inhibitor is present.		

37. *Too much of a good thing.* A simple Michaelis–Menten enzyme, in the absence of any inhibitor, displayed the following kinetic behavior.



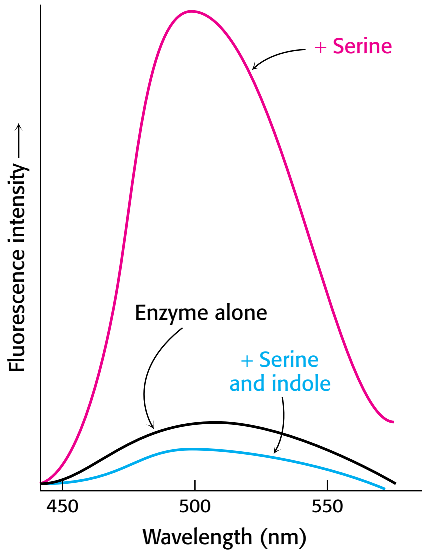
(a) Draw a double-reciprocal plot that corresponds to the velocity-versus-substrate curve.

(b) Suggest a plausible explanation for these kinetic results.

38. *Rate-limiting step.* In the conversion of A into D in the following biochemical pathway, enzymes E_A, E_B, and E_C have the K_M values indicated under each enzyme. If all of the substrates and products are present at a concentration of 10^{-4} M and the enzymes have approximately the same $V_{\max}$, which step will be rate limiting and why?

$$\begin{array}{lccccccc} & A & \underset{E_A}{\rightleftharpoons} & B & \underset{E_B}{\rightleftharpoons} & C & \underset{E_C}{\rightleftharpoons} & D \\ K_M = & & 10^{-2}\,M & & 10^{-4}\,M & & 10^{-4}\,M & \end{array}$$

39. *Colored luminosity* Tryptophan synthetase, a bacterial enzyme that contains a pyridoxal phosphate (PLP) prosthetic group, catalyzes the synthesis of L-tryptophan from L-serine and an indole derivative. The addition of L-serine to the enzyme produces a marked increase in the fluorescence of the PLP group, as the adjoining graph shows. The subsequent addition of indole, the second substrate, reduces this fluorescence to a level even lower than that produced by the enzyme alone. How do these changes in fluorescence support the notion that the enzyme interacts directly with its substrates?



Chapter Integration Problems

40. *Titration experiment.* The effect of pH on the activity of an enzyme was examined. At its active site, the enzyme has an ionizable group that must be negatively charged for substrate binding and catalysis to take place. The ionizable group has a pK_a of 6.0. The substrate is positively charged throughout the pH range of the experiment.

$$\begin{tikzcd} E^- + S^+ \arrow[r, rightleftharpoons] & E^-S^+ \arrow[r] & E^- + P^+ \\ + & & \\ H^+ \arrow[d, rightleftharpoons] & & \\ EH & & \end{tikzcd}$$

(a) Draw the V_0-versus-pH curve when the substrate concentration is much greater than the enzyme K_M.

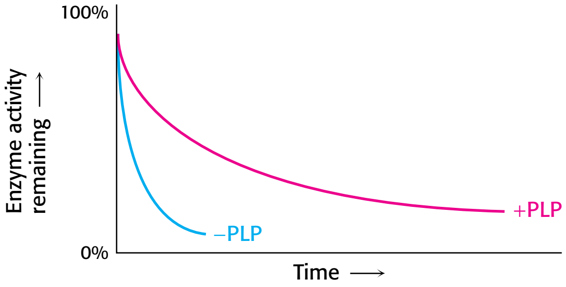
(b) Draw the V_0-versus-pH curve when the substrate concentration is much less than the enzyme K_M.

(c) At which pH will the velocity equal one-half of the maximal velocity attainable under these conditions?

41. *A question of stability.* Pyridoxal phosphate (PLP) is a coenzyme for the enzyme ornithine aminotransferase. The enzyme was purified from cells grown in PLP-deficient media as well as from cells grown in media that contained pyridoxal phosphate. The stability of the two different enzyme preparations was then measured by incubating the enzyme at 37°C for different lengths of time and then assaying for the amount of enzyme activity remaining. The following results were obtained.

(a) Why does the amount of active enzyme decrease with the time of incubation?

(b) Why does the amount of enzyme from the PLP-deficient cells decline more rapidly?

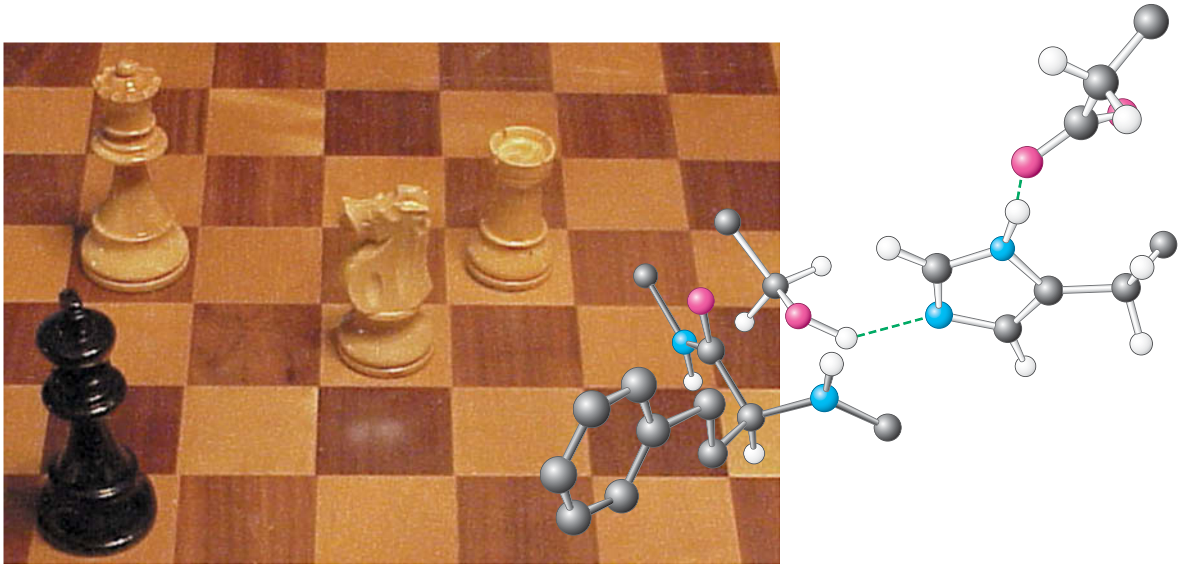


42. *Not just for enzymes*. Kinetics is useful for studying reactions of all types, not just those catalyzed by enzymes. In Chapters 4 and 5, we learned that DNA could be reversibly melted. When melted double-stranded DNA is allowed to renature, the process can be described as consisting of two steps, a slow second order reaction followed by a rapid first order reaction. Explain what is occurring in each step.

CHAPTER

9

Catalytic Strategies



Chess and enzymes have in common the use of strategy, consciously thought out in the game of chess and selected by evolution for the action of an enzyme. The three amino acid residues at the right, denoted by the white bonds, constitute a catalytic triad found in the active site of a class of enzymes that cleave peptide bonds. The substrate, represented by the molecule with the black bonds, is as hopelessly trapped as the king in the photograph of a chess match at the left and is sure to be cleaved. [Photograph courtesy of Wendie Berg.]

What are the sources of the catalytic power and specificity of enzymes? This chapter presents the catalytic strategies used by four classes of enzymes: serine proteases, carbonic anhydrases, restriction endonucleases, and myosins. Each class catalyzes reactions that require the addition of water to a substrate. The mechanisms of these enzymes have been revealed through the use of incisive experimental probes, including the techniques of protein structure determination (Chapter 3) and site-directed mutagenesis (Chapter 5). The mechanisms illustrate many important principles of catalysis. We shall see how these enzymes facilitate the formation of the transition state through the use of binding energy and induced fit as well as additional specific catalytic strategies.

Each of the four classes of enzymes in this chapter illustrates the use of such strategies to solve a different problem. For serine proteases, exemplified by chymotrypsin, the challenge is to promote a reaction that is almost immeasurably slow at neutral pH in the absence of a catalyst. For carbonic anhydrases, the challenge is to achieve a high absolute rate of reaction, suitable for integration with other rapid physiological processes. For restriction endonucleases such as EcoRV, the challenge is to attain a high degree of specificity. Finally, for myosins, the challenge is to utilize the free energy

OUTLINE

9.1 Proteases Facilitate a Fundamentally Difficult Reaction

9.2 Carbonic Anhydrases Make a Fast Reaction Faster

9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work

associated with the hydrolysis of adenosine triphosphate (ATP) to drive other processes. Each of the examples selected is a member of a large protein class. For each of these classes, comparison between class members reveals how enzyme active sites have evolved and been refined. Structural and mechanistic comparisons of enzyme action are thus the sources of insight into the evolutionary history of enzymes. In addition, our knowledge of catalytic strategies has been used to develop practical applications, including potent drugs and specific enzyme inhibitors. Finally, although we shall not consider catalytic RNA molecules explicitly in this chapter, the principles also apply to these catalysts.

A few basic catalytic principles are used by many enzymes

In Chapter 8, we learned that enzymatic catalysis begins with substrate binding. The *binding energy* is the free energy released in the formation of a large number of weak interactions between the enzyme and the substrate. The use of this binding energy is the first common strategy used by enzymes. We can envision this binding energy as serving two purposes: it establishes substrate specificity and increases catalytic efficiency. Only the correct substrate can participate in most or all of the interactions with the enzyme and thus maximize binding energy, accounting for the exquisite substrate specificity exhibited by many enzymes. Furthermore, the full complement of such interactions is formed only when the combination of enzyme and substrate is in the transition state. Thus, interactions between the enzyme and the substrate stabilize the transition state, thereby lowering the free energy of activation. The binding energy can also promote structural changes in both the enzyme and the substrate that facilitate catalysis, a process referred to as *induced fit*.

In addition to the first strategy involving binding energy, enzymes commonly employ one or more of the following four additional strategies to catalyze specific reactions:

1. *Covalent Catalysis*. In covalent catalysis, the active site contains a reactive group, usually a powerful nucleophile, that becomes temporarily covalently attached to a part of the substrate in the course of catalysis. The proteolytic enzyme chymotrypsin provides an excellent example of this strategy (Section 9.1).

2. *General Acid–Base Catalysis*. In general acid–base catalysis, a molecule other than water plays the role of a proton donor or acceptor. Chymotrypsin uses a histidine residue as a base catalyst to enhance the nucleophilic power of serine (Section 9.1), whereas a histidine residue in carbonic anhydrase facilitates the removal of a hydrogen ion from a zinc-bound water molecule to generate hydroxide ion (Section 9.2). For myosins, a phosphate group of the ATP substrate serves as a base to promote its own hydrolysis (Section 9.3).

3. *Catalysis by Approximation*. Many reactions have two distinct substrates, including all four classes of hydrolases considered in detail in this chapter. In such cases, the reaction rate may be considerably enhanced by bringing the two substrates together along a single binding surface on an enzyme. For example, carbonic anhydrase binds carbon dioxide and water in adjacent sites to facilitate their reaction (Section 9.2).

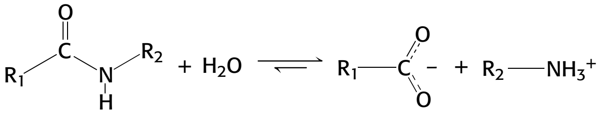
4. *Metal Ion Catalysis*. Metal ions can function catalytically in several ways. For instance, a metal ion may facilitate the formation of nucleophiles such as hydroxide ion by direct coordination. A zinc(II) ion serves this purpose in catalysis by carbonic anhydrase (Section 9.2). Alternatively, a metal ion may serve as an electrophile, stabilizing a negative charge on a reaction intermediate. A magnesium(II) ion plays this role in EcoRV (Section 9.3).

Finally, a metal ion may serve as a bridge between enzyme and substrate, increasing the binding energy and holding the substrate in a conformation appropriate for catalysis. This strategy is used by myosins (Section 9.4) and, indeed, by almost all enzymes that utilize ATP as a substrate.

9.1 Proteases Facilitate a Fundamentally Difficult Reaction

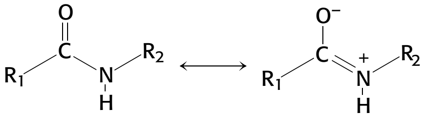
Peptide bond hydrolysis is an important process in living systems (Chapter 23). Proteins that have served their purpose must be degraded so that their constituent amino acids can be recycled for the synthesis of new proteins. Proteins ingested in the diet must be broken down into small peptides and amino acids for absorption in the gut. Furthermore, as described in detail in Chapter 10, proteolytic reactions are important in regulating the activity of certain enzymes and other proteins.

Proteases cleave proteins by a hydrolysis reaction—the addition of a molecule of water to a peptide bond:



Although the hydrolysis of peptide bonds is thermodynamically favorable, such reactions are extremely slow. In the absence of a catalyst, the half-life for the hydrolysis of a typical peptide at neutral pH is estimated to be between 10 and 1000 years. Yet, peptide bonds must be hydrolyzed within milliseconds in some biochemical processes.

The chemical nature of peptide bonds is responsible for their kinetic stability. Specifically, the resonance structure that accounts for the planarity of peptide bonds (Section 2.2) also makes them resistant to hydrolysis. This resonance structure endows them with partial double-bond character:



The carbon–nitrogen bond is strengthened by its double-bond character. Furthermore, the carbonyl carbon atom is less electrophilic and less susceptible to nucleophilic attack than are the carbonyl carbon atoms in more reactive compounds such as carboxylate esters. Consequently, to promote peptide-bond cleavage, an enzyme must facilitate nucleophilic attack at a normally unreactive carbonyl group.

Chymotrypsin possesses a highly reactive serine residue

A number of proteolytic enzymes participate in the breakdown of proteins in the digestive systems of mammals and other organisms. One such enzyme, chymotrypsin, cleaves peptide bonds selectively on the carboxyl-terminal side of the large hydrophobic amino acids such as tryptophan, tyrosine, phenylalanine, and methionine (Figure 9.1). Chymotrypsin is a good example of the use of *covalent catalysis*. The enzyme employs a powerful nucleophile to attack the unreactive carbonyl carbon atom of the substrate. This nucleophile becomes covalently attached to the substrate briefly in the course of catalysis.

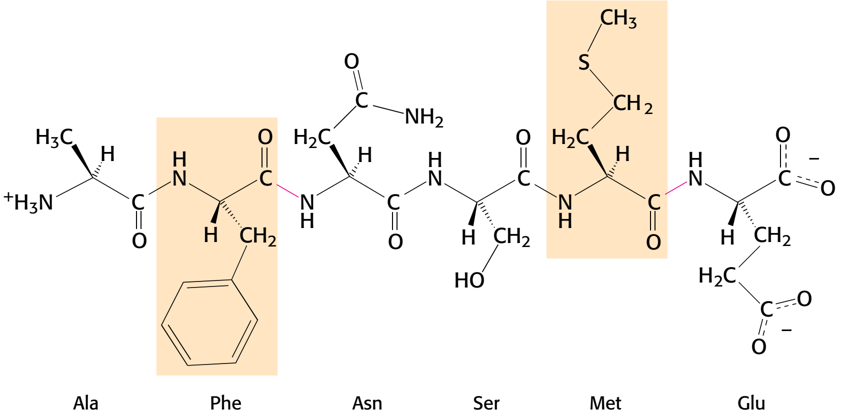


FIGURE 9.1 Specificity of chymotrypsin. Chymotrypsin cleaves proteins on the carboxyl side of aromatic or large hydrophobic amino acids (shaded orange). The likely bonds cleaved by chymotrypsin are indicated in red.

What is the nucleophile that chymotrypsin employs to attack the substrate carbonyl carbon atom? A clue came from the fact that chymotrypsin contains an extraordinarily reactive serine residue. Chymotrypsin molecules treated with organofluorophosphates such as diisopropylphosphofluoridate (DIPF) lost all activity irreversibly (Figure 9.2). Only a single residue, serine 195, was modified. This *chemical modification reaction* suggested that this unusually reactive serine residue plays a central role in the catalytic mechanism of chymotrypsin.

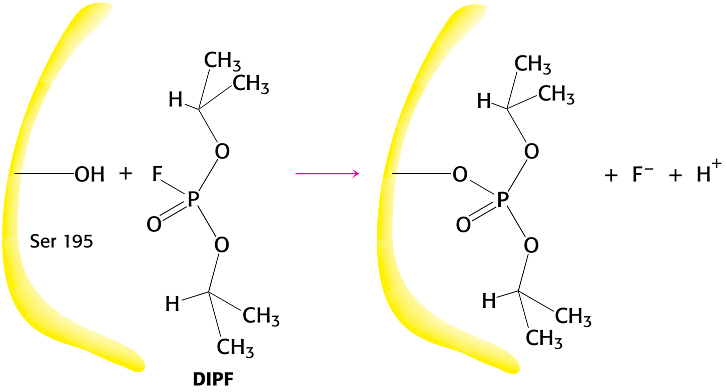


FIGURE 9.2 An unusually reactive serine residue in chymotrypsin. Chymotrypsin is inactivated by treatment with diisopropylphosphofluoridate (DIPF), which reacts only with serine 195 among 28 possible serine residues.

Chymotrypsin action proceeds in two steps linked by a covalently bound intermediate

A study of the kinetics of chymotrypsin provided a second clue to its catalytic mechanism. Enzyme kinetics are often easily monitored by having the enzyme act on a substrate analog that forms a colored product. For chymotrypsin, such a *chromogenic substrate* is *N*-acetyl-L-phenylalanine *p*-nitrophenyl ester. This substrate is an ester rather than an amide, but many proteases will also hydrolyze esters. One of the products formed by chymotrypsin's cleavage of this substrate is *p*-nitrophenolate, which has a yellow color (Figure 9.3). Measurements of the absorbance of light revealed the amount of *p*-nitrophenolate being produced.

Under steady-state conditions, the cleavage of this substrate obeys Michaelis–Menten kinetics with a K_M of 20 μM and a k_{cat} of 77 s^{-1}. The initial phase of the reaction was examined by using the stopped-flow method, which makes it possible to mix enzyme and substrate and monitor the results within a millisecond. This method revealed an initial rapid burst of colored product, followed by its slower formation as the reaction reached

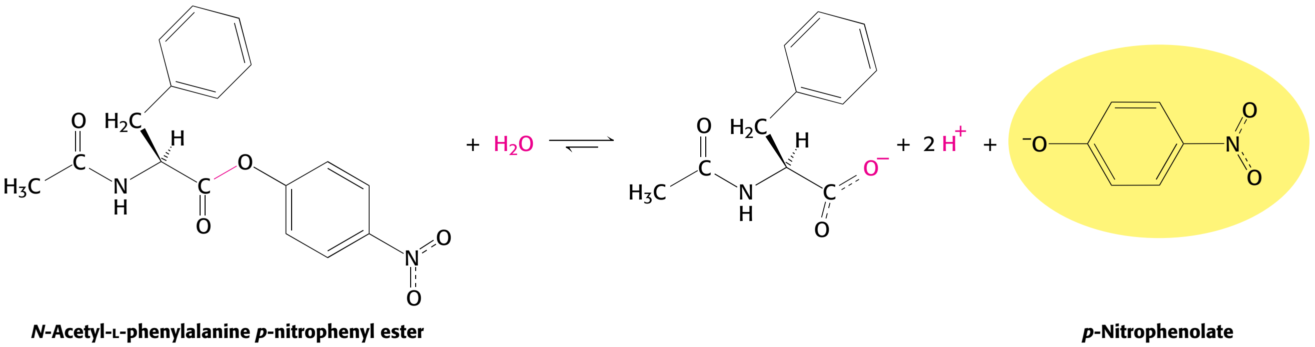


FIGURE 9.3 **Chromogenic substrate.** *N*-Acetyl-L-phenylalanine *p*-nitrophenyl ester yields a yellow product, *p*-nitrophenolate, on cleavage by chymotrypsin. *p*-Nitrophenolate forms by deprotonation of *p*-nitrophenol at pH 7.

the steady state (Figure 9.4). These results suggest that hydrolysis proceeds in two phases. In the first reaction cycle that takes place immediately after mixing, only the first phase must take place before the colored product is released. In subsequent reaction cycles, both phases must take place. Note that the burst is observed because the first phase is substantially more rapid than the second phase for this substrate.

The two phases are explained by the formation of a covalently bound enzyme–substrate intermediate (Figure 9.5). First, the acyl group of the substrate becomes covalently attached to the enzyme as *p*-nitrophenolate (or an amine if the substrate is an amide rather than an ester) is released. The enzyme–acyl group complex is called the *acyl-enzyme intermediate*. Second, the acyl-enzyme intermediate is hydrolyzed to release the carboxylic acid component of the substrate and regenerate the free enzyme. Thus, one molecule of *p*-nitrophenolate is produced rapidly from each enzyme molecule as the acyl-enzyme intermediate is formed. However, it takes longer for the enzyme to be “reset” by the hydrolysis of the acyl-enzyme intermediate, and both phases are required for enzyme turnover.

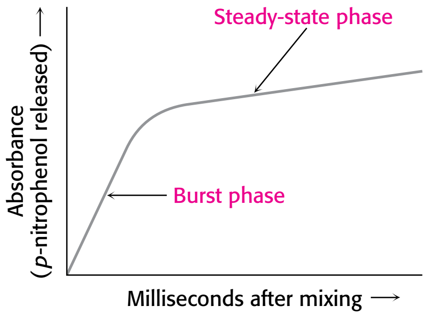


FIGURE 9.4 **Kinetics of chymotrypsin catalysis.** Two phases are evident in the cleaving of *N*-acetyl-L-phenylalanine *p*-nitrophenyl ester by chymotrypsin: a rapid burst phase (pre-steady-state) and a steady-state phase.

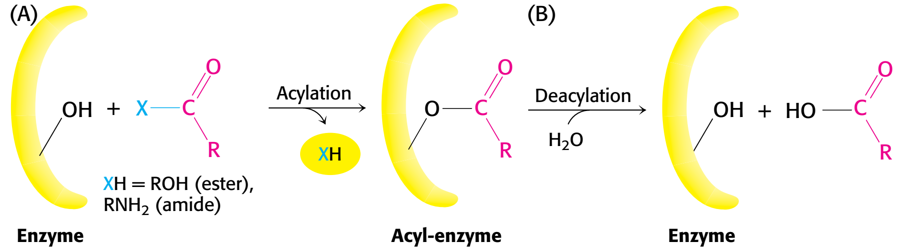


FIGURE 9.5 **Covalent catalysis.** Hydrolysis by chymotrypsin takes place in two phases: (A) acylation to form the acyl-enzyme intermediate followed by (B) deacylation to regenerate the free enzyme.

Serine is part of a catalytic triad that also includes histidine and aspartate

The three-dimensional structure of chymotrypsin revealed that this enzyme is roughly spherical and comprises three polypeptide chains, linked by disulfide bonds. It is synthesized as a single polypeptide, termed *chymotrypsinogen,* which is activated by the proteolytic cleavage of the polypeptide to yield the three chains (Section 10.4). The active site of chymotrypsin, marked by serine 195, lies in a cleft on the surface of the

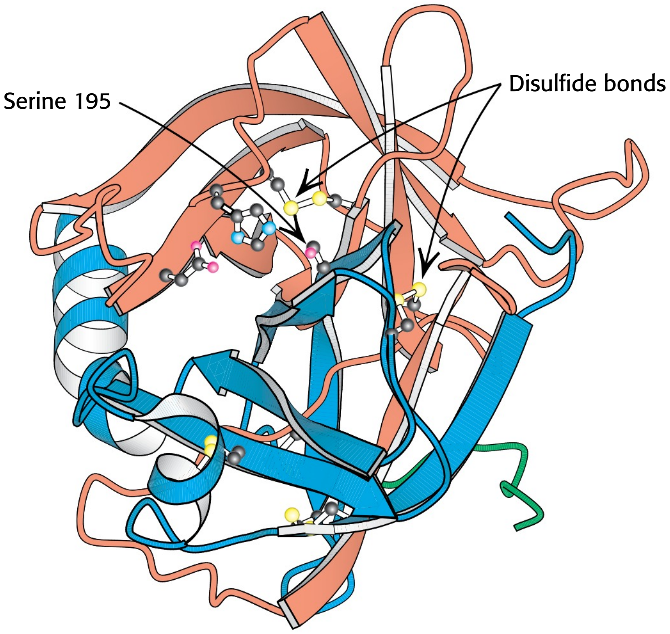


FIGURE 9.6 Location of the active site in chymotrypsin. Chymotrypsin consists of three chains, shown in ribbon form in orange, blue, and green. The side chains of the catalytic triad residues are shown as ball-and-stick representations. *Notice* these side chains, including serine 195, lining the active site in the upper half of the structure. *Also notice* two intrastrand and two interstrand disulfide bonds in various locations throughout the molecule. [Drawn from 1GCT.pdb.]

enzyme (Figure 9.6). The structure of the active site explained the special reactivity of serine 195 (Figure 9.7). The side chain of serine 195 is hydrogen bonded to the imidazole ring of histidine 57. The —NH group of this imidazole ring is, in turn, hydrogen bonded to the carboxylate group of aspartate 102. This constellation of residues is referred to as the *catalytic triad*. How does this arrangement of residues lead to the high reactivity of serine 195? The histidine residue serves to position the serine side chain and to polarize its hydroxyl group so that it is poised for deprotonation. In the presence of the substrate, the histidine residue accepts the proton from the serine 195 hydroxyl group. In doing so, the histidine acts as a general base catalyst. The withdrawal of the proton from the hydroxyl group generates an alkoxide ion, which is a much more powerful nucleophile than is an alcohol. The aspartate residue helps orient the histidine residue and make it a better proton acceptor through hydrogen bonding and electrostatic effects.

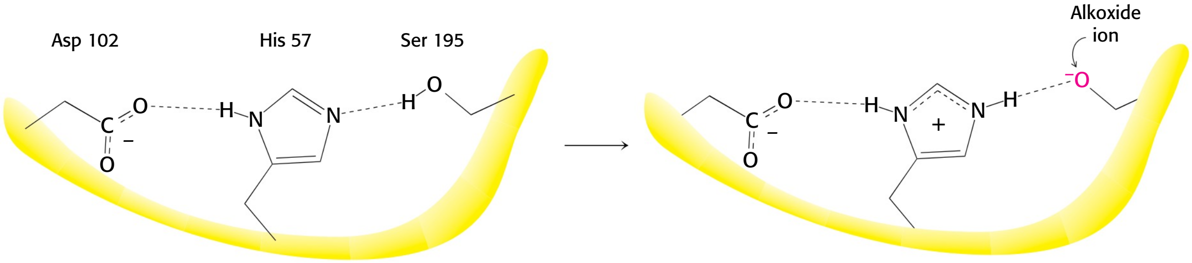


FIGURE 9.7 The catalytic triad. The catalytic triad, shown on the left, converts serine 195 into a potent nucleophile, as illustrated on the right.

These observations suggest a mechanism for peptide hydrolysis (Figure 9.8). After substrate binding (step 1), the reaction begins with the oxygen atom of the side chain of serine 195 making a nucleophilic attack on the carbonyl carbon atom of the target peptide bond (step 2). There are now four atoms bonded to the carbonyl carbon, arranged as a tetrahedron, instead of three atoms in a planar arrangement. This inherently unstable *tetrahedral intermediate* bears a formal negative charge on the oxygen atom derived from the carbonyl group. This charge is stabilized by interactions

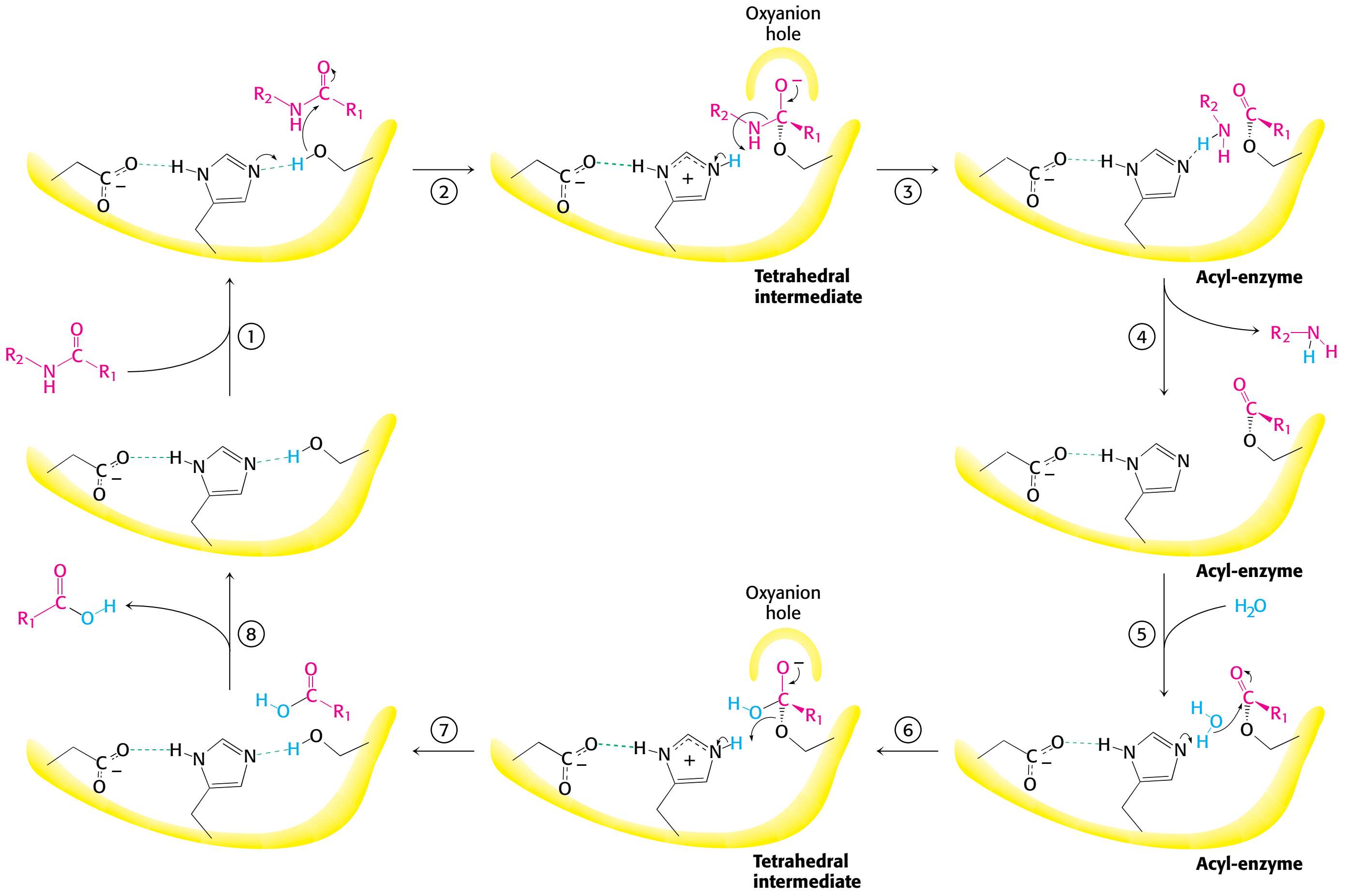


FIGURE 9.8 Peptide hydrolysis by chymotrypsin. The mechanism of peptide hydrolysis illustrates the principles of covalent and acid–base catalysis. The reaction proceeds in eight steps: (1) substrate binding, (2) nucleophilic attack of serine on the peptide carbonyl group, (3) collapse of the tetrahedral intermediate, (4) release of the amine component, (5) water binding, (6) nucleophilic attack of water on the acyl-enzyme intermediate, (7) collapse of the tetrahedral intermediate; and (8) release of the carboxylic acid component. The dashed green lines represent hydrogen bonds.

with NH groups from the protein in a site termed the *oxyanion hole* (Figure 9.9). These interactions also help stabilize the transition state that precedes the formation of the tetrahedral intermediate. This tetrahedral intermediate collapses to generate the acyl-enzyme (step 3). This step is facilitated by the transfer of the proton being held by the positively charged histidine residue to the amino group formed by cleavage of the peptide bond. The amine component is now free to depart from the enzyme (step 4), completing the first stage of the hydrolytic reaction—acylation of the enzyme. Such acyl-enzyme intermediates have even been observed using X-ray crystallography by trapping them through adjustment of conditions such as the nature of the substrate, pH, or temperature.

The next stage—deacylation—begins when a water molecule takes the place occupied earlier by the amine component of the substrate (step 5). The ester group of the acyl-enzyme is now hydrolyzed by a process that essentially repeats steps 2 through 4. Now acting as a general acid catalyst, histidine 57 draws a proton away from the water molecule. The resulting OH^- ion attacks the carbonyl carbon atom of the acyl group, forming a tetrahedral intermediate (step 6). This structure breaks down to form the carboxylic acid product (step 7). Finally, the release of the carboxylic acid product (step 8) readies the enzyme for another round of catalysis.

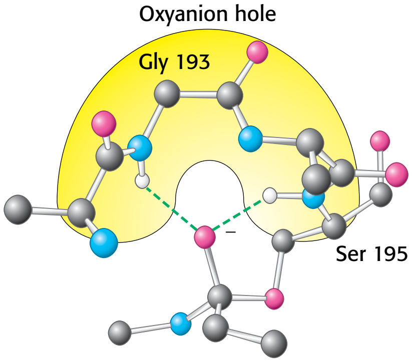


FIGURE 9.9 The oxyanion hole. The structure stabilizes the tetrahedral intermediate of the chymotrypsin reaction. *Notice* that hydrogen bonds (shown in green) link peptide NH groups and the negatively charged oxygen atom of the intermediate.

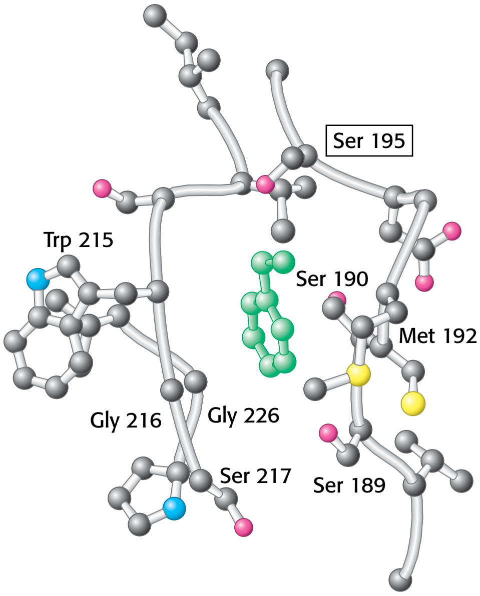


FIGURE 9.10 Specificity pocket of chymotrypsin. *Notice* that this pocket is lined with hydrophobic residues and is deep, favoring the binding of residues with long hydrophobic side chains such as phenylalanine (shown in green). The active-site serine residue (serine 195) is positioned to cleave the peptide backbone between the residue bound in the pocket and the next residue in the sequence. The key amino acids that constitute the binding site are identified.

This mechanism accounts for all characteristics of chymotrypsin action except the observed preference for cleaving the peptide bonds just past residues with large, hydrophobic side chains. Examination of the three-dimensional structure of chymotrypsin with substrate analogs and enzyme inhibitors revealed the presence of a deep hydrophobic pocket, called the S_1 pocket, into which the long, uncharged side chains of residues such as phenylalanine and tryptophan can fit. *The binding of an appropriate side chain into this pocket positions the adjacent peptide bond into the active site for cleavage* (Figure 9.10). The specificity of chymotrypsin depends almost entirely on which amino acid is directly on the amino-terminal side of the peptide bond to be cleaved. Other proteases have more-complex specificity patterns. Such enzymes have additional pockets on their surfaces for the recognition of other residues in the substrate. Residues on the amino-terminal side of the scissile bond (the bond to be cleaved) are labeled P_1, P_2, P_3, and so forth, heading away from the scissile bond (Figure 9.11). Likewise, residues on the carboxyl side of the scissile bond are labeled P_1', P_2', P_3', and so forth. The corresponding sites on the enzyme are referred to as S_1, S_2 or S_1', S_2', and so forth.

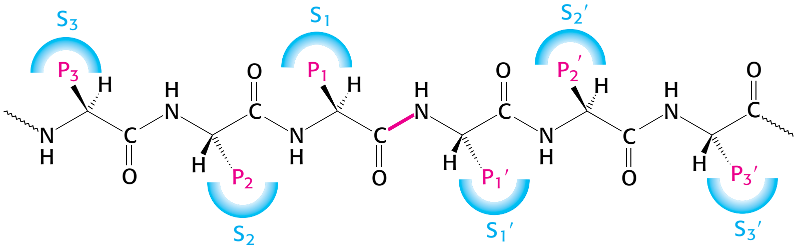


FIGURE 9.11 Specificity nomenclature for protease–substrate interactions. The potential sites of interaction of the substrate with the enzyme are designated P (shown in red), and corresponding binding sites on the enzyme are designated S. The scissile bond (also shown in red) is the reference point.

Catalytic triads are found in other hydrolytic enzymes

Many other peptide-cleaving proteins have subsequently been found to contain catalytic triads similar to that discovered in chymotrypsin. Some, such as trypsin and elastase, are obvious homologs of chymotrypsin. The sequences of these proteins are approximately 40% identical with that of chymotrypsin, and their overall structures are quite similar (Figure 9.12). These proteins operate by mechanisms identical with that of chymotrypsin. However, the three enzymes differ markedly in substrate specificity. Chymotrypsin cleaves at the peptide bond after residues with an aromatic or long nonpolar side chain. Trypsin cleaves at the peptide bond after residues with long, positively charged side chains—namely, arginine and lysine. Elastase cleaves at the peptide bond after amino acids with small side chains—such as alanine and serine. Comparison of the S_1 pockets of these enzymes reveals that *these different specificities are due to small structural differences*. In trypsin, an aspartate residue (Asp 189) is present at the bottom of the S_1 pocket in place of a serine residue in chymotrypsin. The aspartate residue attracts and stabilizes a positively charged arginine or lysine residue in the substrate. In elastase, two residues at the top of the pocket in chymotrypsin and trypsin are replaced by much bulkier valine residues (Val 190 and Val 216). These residues close off the mouth of the pocket so that only small side chains can enter (Figure 9.13).

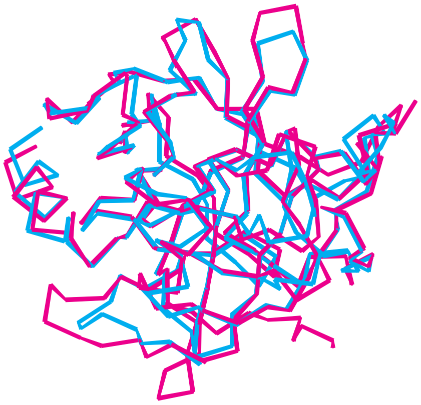


FIGURE 9.12 Structural similarity of trypsin and chymotrypsin. An overlay of the structure of chymotrypsin (red) on that of trypsin (blue) is shown. *Notice* the high degree of similarity. Only α-carbon-atom positions are shown. The mean deviation in position between corresponding α-carbon atoms is 1.7 Å. [Drawn from 5PTP.pdb and 1GCT.pdb.]

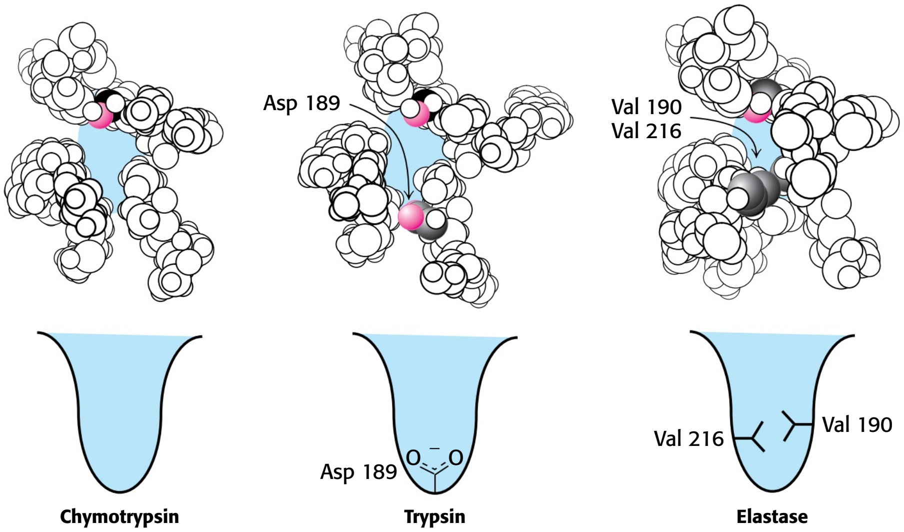


FIGURE 9.13 The S_1 pockets of chymotrypsin, trypsin, and elastase. Certain residues play key roles in determining the specificity of these enzymes. The side chains of these residues, as well as those of the active-site serine residues, are shown in color.

Other members of the chymotrypsin family include a collection of proteins that take part in blood clotting, to be discussed in Chapter 10, as well as the tumor marker protein prostate-specific antigen (PSA). In addition, a wide range of proteases found in bacteria, viruses, and plants belong to this clan.

Other enzymes that are not homologs of chymotrypsin have been found to contain very similar active sites. As noted in Chapter 6, the presence of very similar active sites in these different protein families is a consequence of convergent evolution. Subtilisin, a protease in bacteria such as *Bacillus amyloliquefaciens,* is a particularly well characterized example. The active site of this enzyme includes both the catalytic triad and the oxyanion hole. However, one of the NH groups that forms the oxyanion hole comes from the side chain of an asparagine residue rather than from the peptide backbone (Figure 9.14). Subtilisin is the founding member of another large family of proteases that includes representatives from Archaea, Bacteria, and Eukarya.

Finally, other proteases have been discovered that contain an active-site serine or threonine residue that is activated not by a histidine–aspartate pair but by a primary amino group from the side chain of lysine or by the N-terminal amino group of the polypeptide chain.

Thus, the catalytic triad in proteases has emerged at least three times in the course of evolution. We can conclude that this catalytic strategy

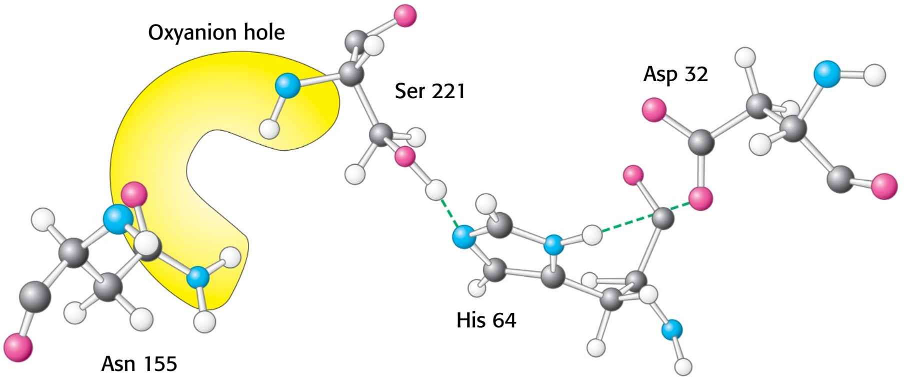


FIGURE 9.14 The catalytic triad and oxyanion hole of subtilisin. *Notice* the two enzyme NH groups (both in the backbone and in the side chain of Asn 155) located in the oxyanion hole. The NH groups will stabilize a negative charge that develops on the peptide bond attacked by nucleophilic serine 221 of the catalytic triad.

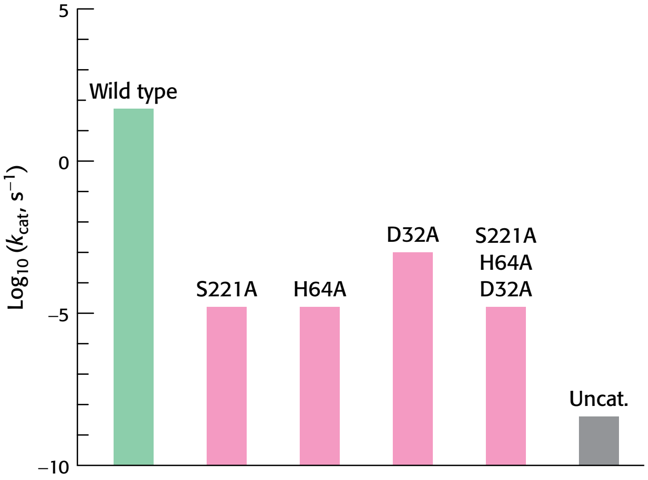


FIGURE 9.15 Site-directed mutagenesis of subtilisin. Residues of the catalytic triad were mutated to alanine, and the activity of the mutated enzyme was measured. Mutations in any component of the catalytic triad cause a dramatic loss of enzyme activity. Note that the activity is displayed on a logarithmic scale. The mutations are identified as follows: the first letter is the one-letter abbreviation for the amino acid being altered; the number identifies the position of the residue in the primary structure; and the second letter is the one-letter abbreviation for the amino acid replacing the original one. Uncat. refers to the estimated rate for the uncatalyzed reaction.

must be an especially effective approach to the hydrolysis of peptides and related bonds.

The catalytic triad has been dissected by site-directed mutagenesis

How can we test the validity of the mechanism proposed for the catalytic triad? One way is to test the contribution of individual amino acid residues to the catalytic power of a protease by using site-directed mutagenesis (Section 5.2). Subtilisin has been extensively studied by this method. Each of the residues within the catalytic triad, consisting of aspartic acid 32, histidine 64, and serine 221, has been individually converted into alanine, and the ability of each mutant enzyme to cleave a model substrate has been examined (Figure 9.15).

As expected, the conversion of active-site serine 221 into alanine dramatically reduced catalytic power; the value of k_{cat} fell to less than *one millionth* of its value for the wild-type enzyme. The value of K_M was essentially unchanged; its increase by no more than a factor of two indicated that substrate continued to bind normally. The mutation of histidine 64 to alanine reduced catalytic power to a similar degree. The conversion of aspartate 32 into alanine reduced catalytic power by less, although the value of k_{cat} still fell to less than 0.005% of its wild-type value. The simultaneous conversion of all three residues into alanine was no more deleterious than the conversion of serine or histidine alone. These observations support the notion that the catalytic triad and, particularly, the serine–histidine pair act together to generate a nucleophile of sufficient power to attack the carbonyl carbon atom of a peptide bond. Despite the reduction in their catalytic power, the mutated enzymes still hydrolyze peptides a thousand times as fast as buffer at pH 8.6.

Site-directed mutagenesis also offered a way to probe the importance of the oxyanion hole for catalysis. The mutation of asparagine 155 to glycine eliminated the side-chain NH group from the oxyanion hole of subtilisin. The elimination of the NH group reduced the value of k_{cat} to 0.2% of its wild-type value but increased the value of K_M by only a factor of two. These observations demonstrate that the NH group of the asparagine residue plays a significant role in stabilizing the tetrahedral intermediate and the transition state leading to it.

Cysteine, aspartyl, and metalloproteases are other major classes of peptide-cleaving enzymes

Not all proteases utilize strategies based on activated serine residues. Classes of proteins have been discovered that employ three alternative

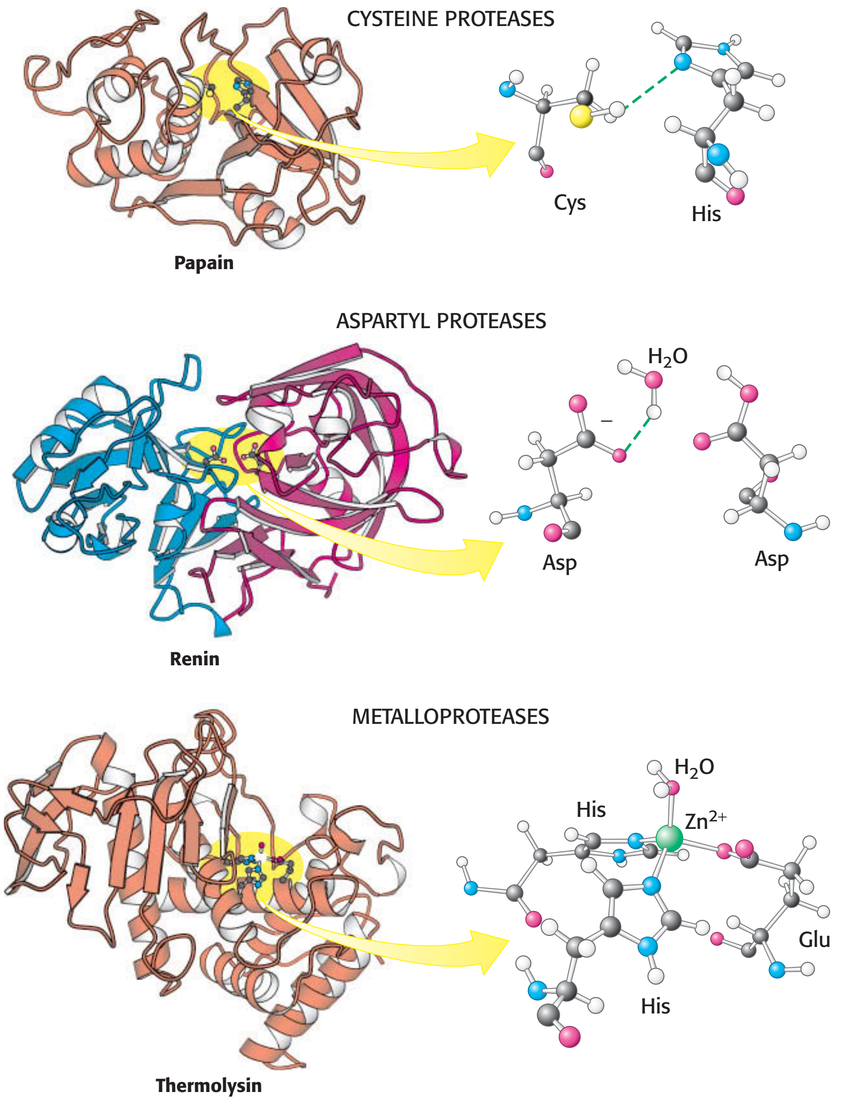
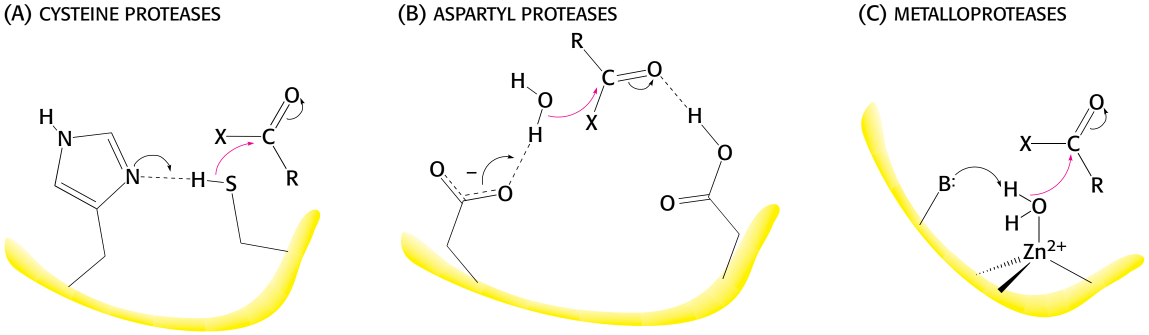

CYSTEINE PROTEASES
Papain
Cys
His
ASPARTYL PROTEASES
Renin
H_2O
−
Asp
Asp
METALLOPROTEASES
Thermolysin
H_2O
His
Zn^{2+}
Glu
His

FIGURE 9.16 Three classes of proteases and their active sites. These examples of a cysteine protease, an aspartyl protease, and a metalloprotease use a histidine-activated cysteine residue, an aspartate-activated water molecule, and a metal-activated water molecule, respectively, as the nucleophile. The two halves of renin are in blue and red to highlight the approximate twofold symmetry of aspartyl proteases. *Notice* how different these active sites are despite the similarity in the reactions they catalyze. [Drawn from 1PPN.pdb.; 1HRN. pdb; 1LND.pdb.]

approaches to peptide-bond hydrolysis (Figure 9.16). These classes are the (1) cysteine proteases, (2) aspartyl proteases, and (3) metalloproteases. In each case, the strategy is to generate a nucleophile that attacks the peptide carbonyl group (Figure 9.17).

The strategy used by the *cysteine proteases* is most similar to that used by the chymotrypsin family. In these enzymes, a cysteine residue, activated by a histidine residue, plays the role of the nucleophile that attacks the peptide bond (Figure 9.17) in a manner quite analogous to that of the serine residue in serine proteases. Because the sulfur atom in cysteine is inherently a better

FIGURE 9.17 The activation strategies for three classes of proteases. The peptide carbonyl group is attacked by (A) a histidine-activated cysteine in the cysteine proteases, (B) an aspartate-activated water molecule in the aspartyl proteases, and (C) a metal-activated water molecule in the metalloproteases. For the metalloproteases, the letter B represents a base (often glutamate) that helps deprotonate the metal-bound water.


(A) CYSTEINE PROTEASES
H
N
N
H
S
X
C
O
R
(B) ASPARTYL PROTEASES
R
C
O
H
O
H
X
H
O
O
−
O
O
(C) METALLOPROTEASES
O
X
C
R
B:
H
O
H
Zn^{2+}

nucleophile than is the oxygen atom in serine, cysteine proteases appear to require only this histidine residue in addition to cysteine and not the full catalytic triad. A well-studied example of these proteins is papain, an enzyme purified from the fruit of the papaya. Mammalian proteases homologous to papain have been discovered, most notably the cathepsins, proteins having a role in the immune system and other systems. The cysteine-based active site arose independently at least twice in the course of evolution; the caspases, enzymes that play a major role in apoptosis (a genetically programmed cell death pathway), have active sites similar to that of papain, but their overall structures are unrelated.

The second class comprises the *aspartyl proteases.* The central feature of the active sites is a pair of aspartic acid residues that act together to allow a water molecule to attack the peptide bond. One aspartic acid residue (in its deprotonated form) activates the attacking water molecule by poising it for deprotonation. The other aspartic acid residue (in its protonated form) polarizes the peptide carbonyl group so that it is more susceptible to attack (Figure 9.17). Members of this class include renin, an enzyme involved in the regulation of blood pressure, and the digestive enzyme pepsin. These proteins possess approximate twofold symmetry. A likely scenario is that two copies of a gene for the ancestral enzyme fused to form a single gene that encoded a single-chain enzyme. Each copy of the gene would have contributed an aspartate residue to the active site. The individual chains are now joined to make a single chain in most aspartyl proteases, whereas the proteases present in human immunodeficiency virus (HIV) and other retroviruses comprise dimers of identical chains (Figure 9.18). This observation is consistent with the idea that larger aspartyl proteases may have evolved by fusion of separate subunits.

The *metalloproteases* constitute the final major class of peptide-cleaving enzymes. The active site of such a protein contains a bound metal ion, almost always zinc, that activates a water molecule to act as a nucleophile to attack the peptide carbonyl group. The bacterial enzyme thermolysin and the digestive enzyme carboxypeptidase A are classic examples of the zinc proteases. Thermolysin, but not carboxypeptidase A, is a member of a large and diverse family of homologous zinc proteases that includes the matrix metalloproteases, enzymes that catalyze the reactions in tissue remodeling and degradation.

In each of these three classes of enzymes, the active site includes features that act to (1) activate a water molecule or another nucleophile, (2) polarize the peptide carbonyl group, and (3) stabilize a tetrahedral intermediate (Figure 9.17).

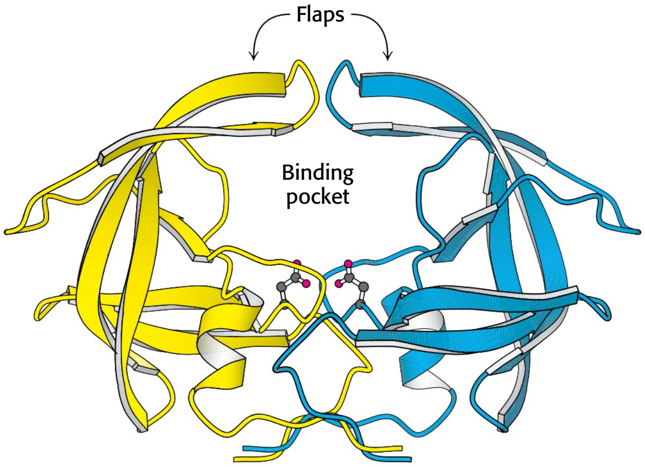


FIGURE 9.18 HIV protease, a dimeric aspartyl protease. The protease is a dimer of identical subunits, shown in blue and yellow, consisting of 99 amino acids each. *Notice* the placement of active-site aspartic acid residues, one from each chain, which are shown as ball-and-stick structures. The flaps will close down on the binding pocket after substrate has been bound. [Drawn from 3PHV.pdb.]

Protease inhibitors are important drugs

Several important drugs are protease inhibitors. For example, captopril, used to regulate blood pressure, is one of many inhibitors of the angiotensin-converting enzyme (ACE), a metalloprotease. Indinavir (Crixivan), retrovir, and more than 20 other compounds used in the treatment of AIDS are inhibitors of HIV protease (Figure 9.18), an aspartyl protease. HIV protease cleaves multidomain viral proteins into their active forms; blocking this process completely prevents the virus from being infectious. HIV protease inhibitors, in combination with inhibitors of other key HIV enzymes, dramatically reduced deaths due to AIDS, assuming that the cost of the treatment could be covered (Figure 36.21). In many cases, these drugs have converted AIDS from a death sentence to a treatable chronic disease.

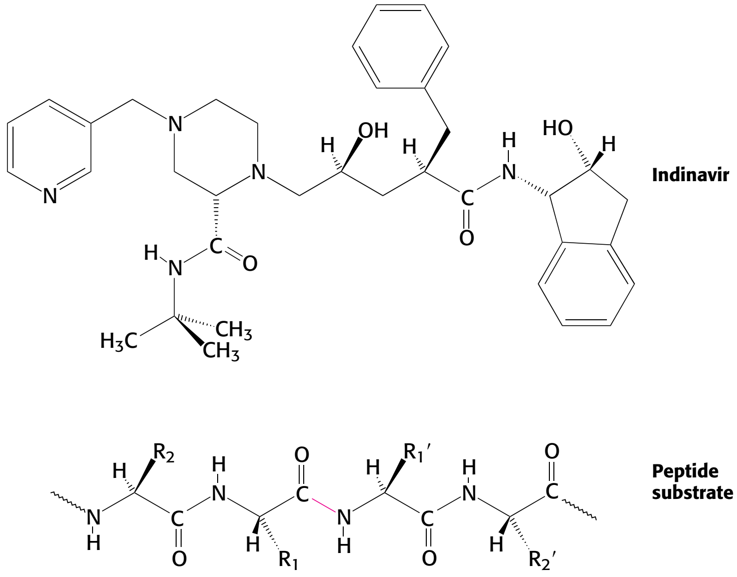


FIGURE 9.19 Indinavir, an HIV protease inhibitor. The structure of indinavir (Crixivan) is shown in comparison with that of a peptide substrate of HIV protease. The scissile bond in the substrate is highlighted in red.

Indinavir resembles the peptide substrate of the HIV protease. Indinavir is constructed around an alcohol that mimics the tetrahedral intermediate; other groups are present to bind into the S_2, S_1, S_1', and S_2' recognition sites on the enzyme (Figure 9.19). X-ray crystallographic studies revealed that, in the active site, indinavir adopts a confirmation that approximates the twofold symmetry of the enzyme (Figure 9.20).

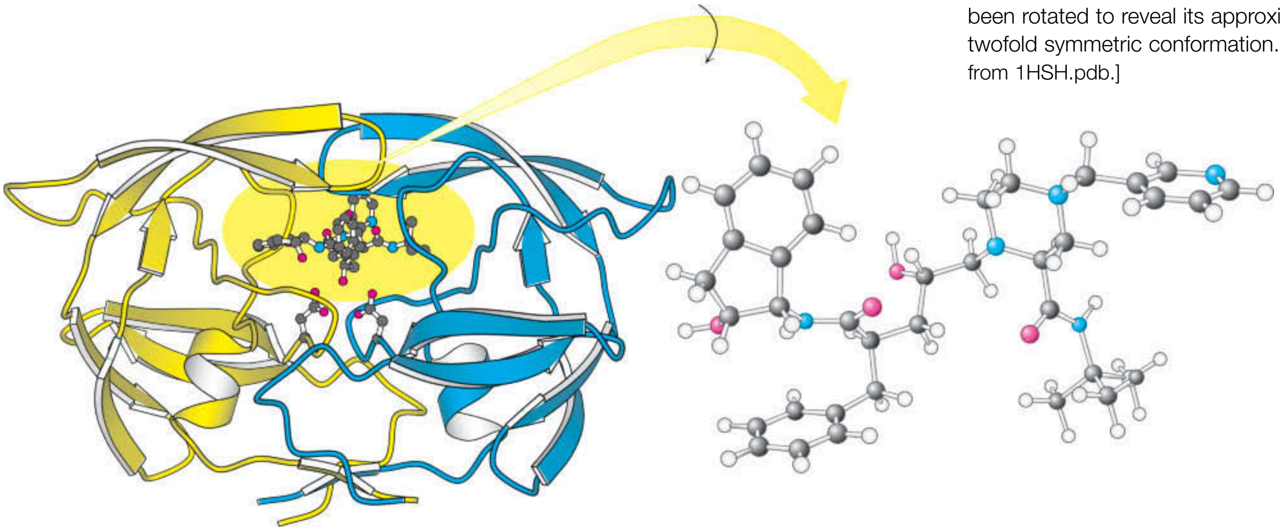
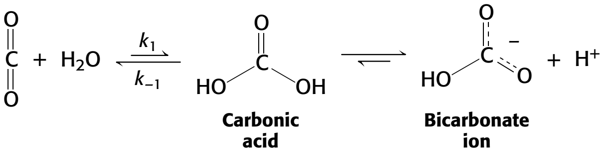


FIGURE 9.20 HIV protease–indinavir complex. (Left) The HIV protease is shown with the inhibitor indinavir bound at the active site. *Notice* the twofold symmetry of the enzyme structure. (Right) The drug has been rotated to reveal its approximately twofold symmetric conformation. [Drawn from 1HSH.pdb.]

The active site of HIV protease is covered by two flexible flaps that fold down on top of the bound inhibitor. The OH group of the central alcohol interacts with the two aspartate residues of the active site. In addition, two carbonyl groups of the inhibitor are hydrogen bonded to a water molecule (not shown in Figure 9.20), which, in turn, is hydrogen bonded to a peptide NH group in each of the flaps. This interaction of the inhibitor with water and the enzyme is not possible within cellular aspartyl proteases such as renin. Thus, the interaction may contribute to the specificity of indinavir for HIV protease. To prevent side effects, protease inhibitors used as drugs must be specific for one enzyme without inhibiting other proteins within the body.

9.2 Carbonic Anhydrases Make a Fast Reaction Faster

Carbon dioxide is a major end product of aerobic metabolism. In mammals, this carbon dioxide is released into the blood and transported to the lungs for exhalation. While in the red blood cells, carbon dioxide reacts with water (Section 7.3). The product of this reaction is a moderately strong acid, carbonic acid ($pK_a = 3.5$), which is converted into bicarbonate ion (HCO_3^-) on the loss of a proton.



Even in the absence of a catalyst, this hydration reaction proceeds at a moderately fast pace. At 37°C near neutral pH, the second-order rate constant k_1 is 0.0027 $M^{-1}\ s^{-1}$. This value corresponds to an effective first-order rate constant of 0.15 s^{-1} in water ($[H_2O] = 55.5$ M). The reverse reaction, the dehydration of HCO_3^-, is even more rapid, with a rate constant of $k_{-1} = 50\ s^{-1}$. These rate constants correspond to an equilibrium constant of $K_1 = 5.4 \times 10^{-5}$ and a ratio of $[CO_2]$ to $[H_2CO_3]$ of 340 : 1 at equilibrium.

Carbon dioxide hydration and HCO_3^- dehydration are often coupled to rapid processes, particularly transport processes. Thus, almost all organisms contain enzymes, referred to as *carbonic anhydrases,* that increase the rate of reaction beyond the already reasonable spontaneous rate. For example, carbonic anhydrases dehydrate HCO_3^- in the blood to form CO_2 for exhalation as the blood passes through the lungs. Conversely, they convert CO_2 into HCO_3^- to generate the aqueous humor of the eye and other secretions. Furthermore, both CO_2 and HCO_3^- are substrates and products for a variety of enzymes, and the rapid interconversion of these species may be necessary to ensure appropriate substrate levels. So important are these enzymes in human beings that mutations in some carbonic anhydrases have been found to be associated with osteopetrosis (excessive formation of dense bones accompanied by anemia) and mental retardation.

Carbonic anhydrases accelerate CO_2 hydration dramatically. The most-active enzymes hydrate CO_2 at rates as high as $k_{cat} = 10^6\ s^{-1}$, or a million times a second per enzyme molecule. Fundamental physical processes such as diffusion and proton transfer ordinarily limit the rate of hydration, and so the enzymes employ special strategies to attain such prodigious rates.

Carbonic anhydrase contains a bound zinc ion essential for catalytic activity

Less than 10 years after the discovery of carbonic anhydrase in 1932, this enzyme was found to contain a bound zinc ion. Moreover, the zinc ion appeared to be necessary for catalytic activity. This discovery, remarkable at the time, made carbonic anhydrase the first known zinc-containing enzyme. At present, hundreds of enzymes are known to contain zinc. In fact, more than one-third of all enzymes either contain bound metal ions or require the addition of such ions for activity. Metal ions have several properties that increase chemical reactivity: their positive charges, their ability to form strong yet kinetically labile bonds, and, in some cases, their capacity to be stable in more than one oxidation state. The chemical reactivity of metal ions explains why catalytic strategies that employ metal ions have been adopted throughout evolution.

X-ray crystallographic studies have supplied the most-detailed and direct information about the zinc site in carbonic anhydrase. At least seven carbonic anhydrases, each with its own gene, are present in human beings. They are all clearly homologous, as revealed by substantial sequence identity. Carbonic anhydrase II, a major protein component of red blood cells, has been the most extensively studied (Figure 9.21). It is also one of the most active carbonic anhydrases.

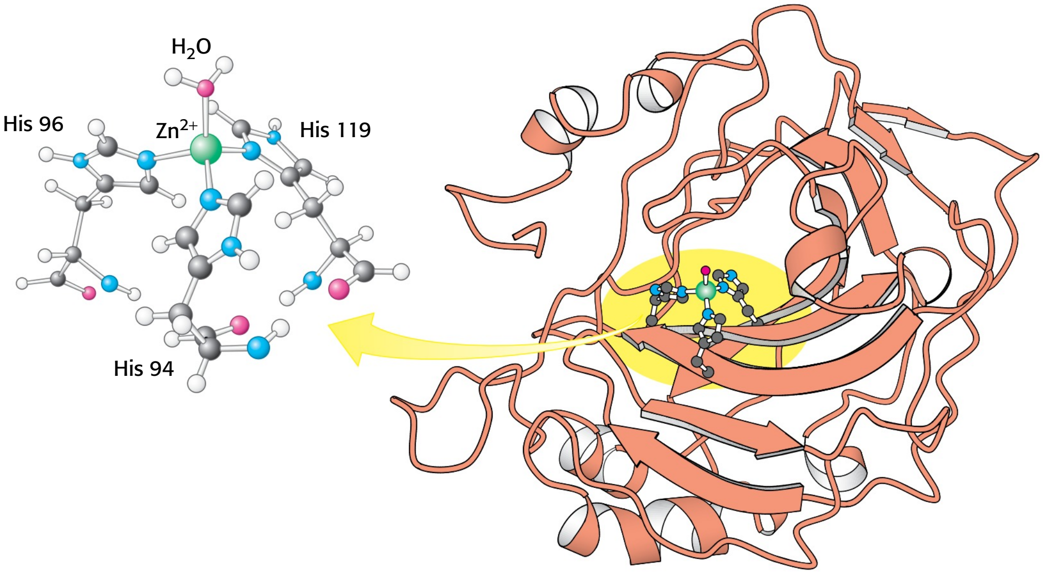


FIGURE 9.21 The structure of human carbonic anhydrase II and its zinc site. (Left) *Notice* that the zinc ion is bound to the imidazole rings of three histidine residues as well as to a water molecule. (Right) *Notice* the location of the zinc site in a cleft near the center of the enzyme. [Drawn from 1CA2.pdb.]

Zinc is found only in the +2 state in biological systems. A zinc atom is essentially always bound to four or more ligands; in carbonic anhydrase, three coordination sites are occupied by the imidazole rings of three histidine residues and an additional coordination site is occupied by a water molecule (or hydroxide ion, depending on pH). Because the molecules occupying the coordination sites are neutral, the overall charge on the $Zn(His)_3$ unit remains +2.

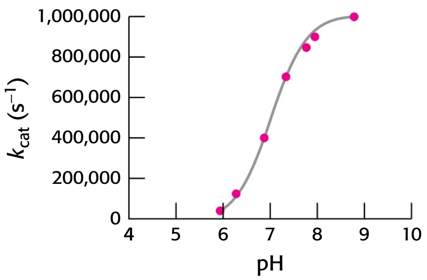


FIGURE 9.22 Effect of pH on carbonic anhydrase activity. Changes in pH alter the rate of carbon dioxide hydration catalyzed by carbonic anhydrase II. The enzyme is maximally active at high pH.

Catalysis entails zinc activation of a water molecule

How does this zinc complex facilitate carbon dioxide hydration? A major clue comes from the pH profile of enzymatically catalyzed carbon dioxide hydration (Figure 9.22).

At pH 8, the reaction proceeds near its maximal rate. As the pH decreases, the rate of the reaction drops. The midpoint of this transition is

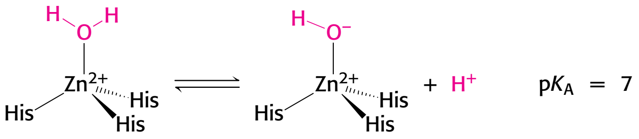


FIGURE 9.23 The pK_a of zinc-bound water. Binding to zinc lowers the pK_a of water from 15.7 to 7.

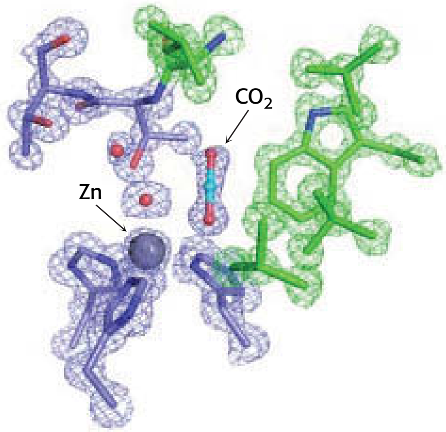


FIGURE 9.24 Carbon dioxide binding site. Crystals of carbonic anhydrase were exposed to carbon dioxide gas at high pressure and low temperature, and x-ray diffraction data were collected. The electron density for carbon dioxide, clearly visible adjacent to the zinc and its bound water, reveals the carbon dioxide binding site. [Information from J. F. Domsic, et al., *J. Biol. Chem.* 283:30766–30771, 2008.]

near pH 7, suggesting that a group that loses a proton at pH 7 (pK_a = 7) plays an important role in the activity of carbonic anhydrase. Moreover, the curve suggests that the deprotonated (high pH) form of this group participates more effectively in catalysis. Although some amino acids, notably histidine, have pK_a values near 7, *a variety of evidence suggests that the group responsible for this transition is not an amino acid but is the zinc-bound water molecule.*

The binding of a water molecule to the positively charged zinc center reduces the pK_a of the water molecule from 15.7 to 7 (Figure 9.23). With the pK_a lowered, the water molecule can more easily lose a proton at neutral pH, generating a substantial concentration of hydroxide ion (bound to the zinc atom). A zinc-bound hydroxide ion (OH^-) is a potent nucleophile able to attack carbon dioxide much more readily than water does. Adjacent to the zinc site, carbonic anhydrase also possesses a hydrophobic patch that serves as a binding site for carbon dioxide (Figure 9.24). Based on these observations, a simple mechanism for carbon dioxide hydration can be proposed (Figure 9.25):

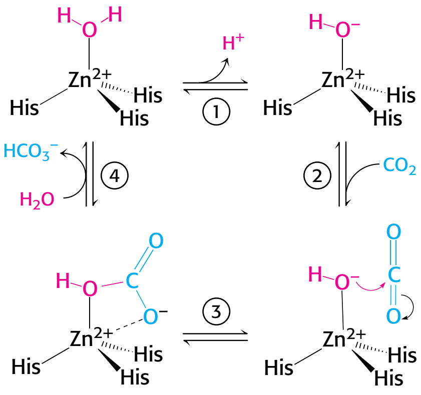


FIGURE 9.25 Mechanism of carbonic anhydrase. The zinc-bound hydroxide mechanism for the hydration of carbon dioxide reveals one aspect of metal ion catalysis. The reaction proceeds in four steps: (1) water deprotonation, (2) carbon dioxide binding, (3) nucleophilic attack by hydroxide on carbon dioxide, and (4) displacement of bicarbonate ion by water.

1. The zinc ion facilitates the release of a proton from a water molecule, which generates a hydroxide ion.

2. The carbon dioxide substrate binds to the enzyme's active site and is positioned to react with the hydroxide ion.

3. The hydroxide ion attacks the carbon dioxide, converting it into bicarbonate ion, HCO_3^-.

4. The catalytic site is regenerated with the release of HCO_3^- and the binding of another molecule of water.

Thus, the binding of a water molecule to the zinc ion favors the formation of the transition state by facilitating proton release and by positioning the water molecule to be in close proximity to the other reactant.

Studies of a *synthetic analog model system* provide evidence for the mechanism's plausibility. A simple synthetic ligand binds zinc through four

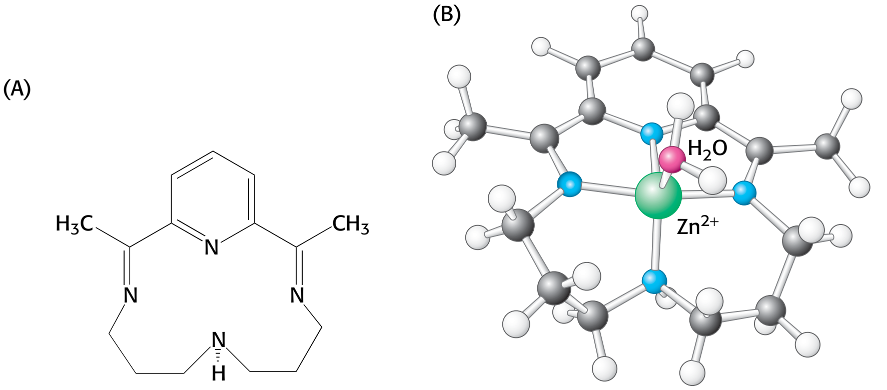


FIGURE 9.26 A synthetic analog model system for carbonic anhydrase. (A) An organic compound, capable of binding zinc, was synthesized as a model for carbonic anhydrase. The zinc complex of this ligand accelerates the hydration of carbon dioxide more than 100-fold under appropriate conditions. (B) The structure of the presumed active complex showing zinc bound to the ligand and to one water molecule.

nitrogen atoms (compared with three histidine nitrogen atoms in the enzyme), as shown in Figure 9.26. One water molecule remains bound to the zinc ion in the complex. Direct measurements reveal that this water molecule has a pK_a value of 8.7, not as low as the value for the water molecule in carbonic anhydrase but substantially lower than the value for free water. At pH 9.2, this complex accelerates the hydration of carbon dioxide more than 100-fold. Although its rate of catalysis is much less efficient than catalysis by carbonic anhydrase, the model system strongly suggests that the zinc-bound hydroxide mechanism is likely to be correct. Carbonic anhydrases have evolved to employ the reactivity intrinsic to a zinc-bound hydroxide ion as a potent catalyst.

A proton shuttle facilitates rapid regeneration of the active form of the enzyme

As noted earlier, some carbonic anhydrases can hydrate carbon dioxide at rates as high as a million times a second (10^6 s^{-1}). The magnitude of this rate can be understood from the following observations. In the first step of a carbon dioxide hydration reaction, the zinc-bound water molecule must lose a proton to regenerate the active form of the enzyme (Figure 9.27). The rate of the reverse reaction, the protonation of the zinc-bound hydroxide ion, is limited by the rate of proton diffusion. Protons diffuse very rapidly with second-order rate constants near 10^{11} M^{-1} s^{-1}. Thus, the backward rate constant k_{-1} must be less than 10^{11} M^{-1} s^{-1}. Because the equilibrium constant K is equal to k_1/k_{-1}, the forward rate constant is given by $k_1 = K \cdot k_{-1}$. Thus, if $k_{-1} \leq 10^{11}$ M^{-1} s^{-1} and $K = 10^{-7}$ M (because p$K_a = 7$), then k_1 must be less than or equal to 10^4 s^{-1}. In other words, the rate of proton diffusion limits the rate of proton release to less than 10^4 s^{-1} for a group with p$K_a = 7$. However, if carbon dioxide is hydrated at a rate of 10^6 s^{-1}, then every step in the mechanism (Figure 9.25) must take place at least this fast. How is this apparent paradox resolved?

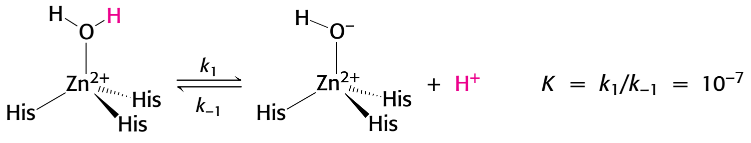
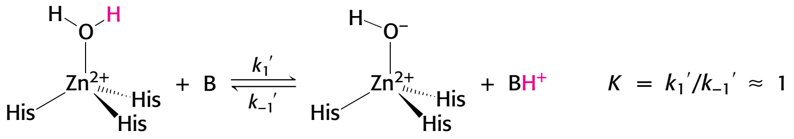


FIGURE 9.27 Kinetics of water deprotonation. The kinetics of deprotonation and protonation of the zinc-bound water molecule in carbonic anhydrase.

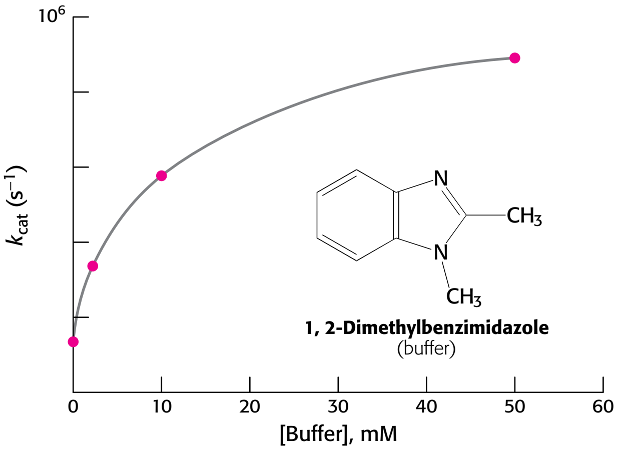
The answer became clear with the realization that *the highest rates of carbon dioxide hydration require the presence of buffer, suggesting that the buffer components participate in the reaction.* The buffer can bind or release protons. The advantage is that, whereas the concentrations of

FIGURE 9.28 The effect of buffer on deprotonation. The deprotonation of the zinc-bound water molecule in carbonic anhydrase is aided by buffer component B.



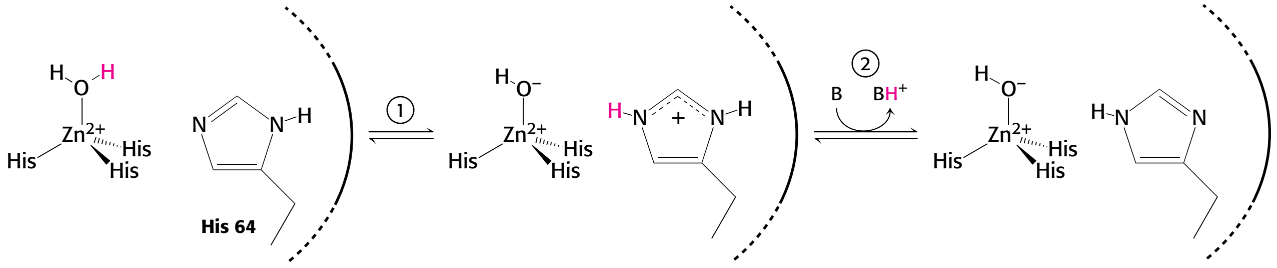
protons and hydroxide ions are limited to 10^{-7} M at neutral pH, the concentration of buffer components can be much higher, of the order of several millimolar. If the buffer component BH^+ has a pK_a of 7 (matching that for the zinc-bound water molecule), then the equilibrium constant for the reaction in Figure 9.28 is 1. The rate of proton abstraction is given by $k_1' \cdot [B]$. The second-order rate constants k_1' and k_{-1}' will be limited by buffer diffusion to values less than approximately $10^9\ M^{-1}\ s^{-1}$. Thus, buffer concentrations greater than $[B] = 10^{-3}$ M (or 1 mM) may be high enough to support carbon dioxide hydration rates of $10^6\ M^{-1}\ s^{-1}$ because $k_1' \cdot [B] = (10^9\ M^{-1}s^{-1}) \cdot (10^{-3}M) = 10^6\ s^{-1}$. The prediction that the rate increases with increasing buffer concentration has been confirmed experimentally (Figure 9.29).

FIGURE 9.29 The effect of buffer concentration on the rate of carbon dioxide hydration. The rate of carbon dioxide hydration increases with the concentration of the buffer 1,2-dimethylbenzimidazole. The buffer enables the enzyme to achieve its high catalytic rates.



The molecular components of many buffers are too large to reach the active site of carbonic anhydrase. Carbonic anhydrase II has evolved a *proton shuttle* to allow buffer components to participate in the reaction from solution. The primary component of this shuttle is histidine 64. This residue transfers protons from the zinc-bound water molecule to the protein surface and then to the buffer (Figure 9.30). Thus, catalytic function has been enhanced through the evolution of an apparatus for controlling proton transfer from and to the active site. Because protons participate in many biochemical reactions, the manipulation of the proton inventory within active sites is crucial to the function of many enzymes and explains the prominence of acid–base catalysis.

FIGURE 9.30 Histidine proton shuttle. (1) Histidine 64 abstracts a proton from the zinc-bound water molecule, generating a nucleophilic hydroxide ion and a protonated histidine. (2) The buffer (B) removes a proton from the histidine, regenerating the unprotonated form.



9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions

We next consider a hydrolytic reaction that results in the cleavage of DNA. Bacteria and archaea have evolved mechanisms to protect themselves from viral infections. Many viruses inject their DNA genomes into cells; once inside, the viral DNA hijacks the cell's machinery to drive the production of viral proteins and, eventually, of progeny virus. Often, a viral infection results in the death of the host cell. A major protective strategy for the host is to use *restriction endonucleases* (restriction enzymes) to degrade the viral DNA on its introduction into a cell. These enzymes recognize particular base sequences, called *recognition sequences* or *recognition sites,* in their target DNA and cleave that DNA at defined positions. We have already considered the utility of these important enzymes for dissecting genes and genomes (Section 5.1). The most well-studied class of restriction enzymes comprises the type II restriction enzymes, which cleave DNA *within* their recognition sequences. Other types of restriction enzymes cleave DNA at positions somewhat distant from their recognition sites.

Restriction endonucleases must show tremendous specificity at two levels. First, they must not degrade host DNA containing the recognition sequences. Second, they must cleave only DNA molecules that contain recognition sites (hereafter referred to as *cognate DNA*) without cleaving DNA molecules that lack these sites. How do these enzymes manage to degrade viral DNA while sparing their own? In *E. coli,* the restriction endonuclease EcoRV cleaves double-stranded viral DNA molecules that contain the sequence 5′-GATATC-3′ but leaves intact host DNA containing hundreds of such sequences. We shall return to the strategy by which host cells protect their own DNA at the end of this section.

Restriction enzymes must cleave DNA only at recognition sites, without cleaving at other sites. Suppose that a recognition sequence is six base pairs long. Because there are 4^6, or 4096, sequences having six base pairs, the concentration of sites that must not be cleaved will be approximately 4000-fold higher than the concentration of sites that should be cleaved. Thus, to keep from damaging host-cell DNA, restriction enzymes must cleave cognate DNA molecules much more than 4000 times as efficiently as they cleave nonspecific sites. We shall return to the mechanism used to achieve the necessary high specificity after considering the chemistry of the cleavage process.

Cleavage is by in-line displacement of 3′-oxygen from phosphorus by magnesium-activated water

A restriction endonuclease catalyzes the hydrolysis of the phosphodiester backbone of DNA. Specifically, the bond between the 3′-oxygen atom and the phosphorus atom is broken. The products of this reaction are DNA strands with a free 3′-hydroxyl group and a 5′-phosphoryl group at the cleavage site (Figure 9.31). This reaction proceeds by nucleophilic attack at the phosphorus atom. We will consider two alternative mechanisms,

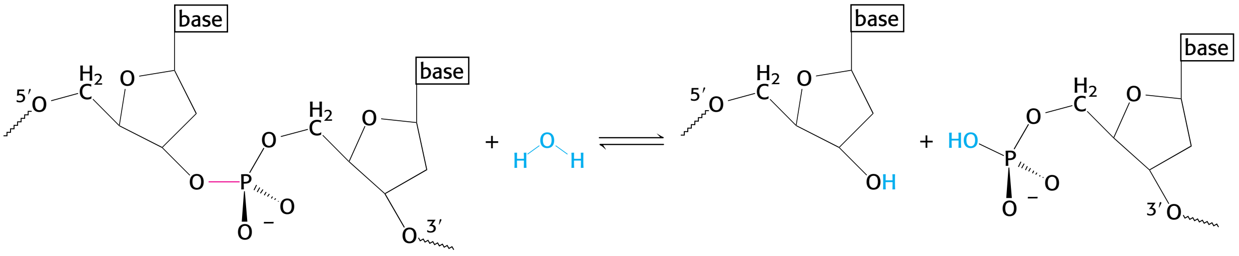
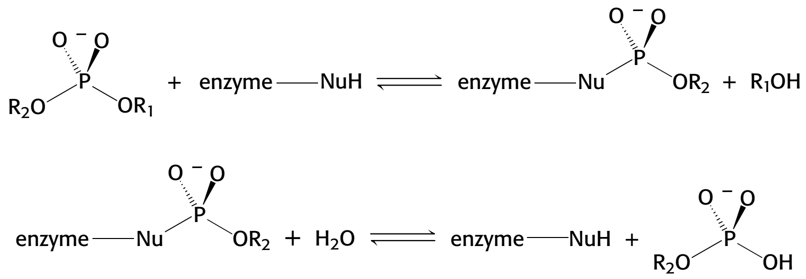


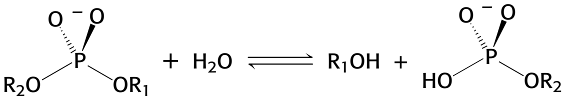
FIGURE 9.31 Hydrolysis of a phosphodiester bond. All restriction enzymes catalyze the hydrolysis of DNA phosphodiester bonds, leaving a phosphoryl group attached to the 5' end. The bond that is cleaved is shown in red.

suggested by analogy with the proteases. The restriction endonuclease might cleave DNA by mechanism 1 through a covalent intermediate, employing a potent nucleophile (Nu), or by mechanism 2 through direct hydrolysis:

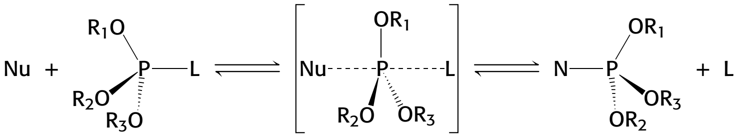
Mechanism 1 (covalent intermediate)



Mechanism 2 (direct hydrolysis)



Each mechanism postulates a different nucleophile to attack the phosphorus atom. In either case, each reaction takes place by *in-line displacement:*



The incoming nucleophile attacks the phosphorus atom, and a pentacoordinate transition state is formed. This species has a trigonal bipyramidal geometry centered at the phosphorus atom, with the incoming nucleophile at one apex of the two pyramids and the group that is displaced (the leaving group, L) at the other apex. Note that the displacement inverts the stereochemical conformation at the tetrahedral phosphorous atom, analogous to the interconversion of the R and S configurations around a tetrahedral carbon center (Section 2.1).

The two mechanisms differ in the number of times that the displacement takes place in the course of the reaction. In the first type of mechanism, a nucleophile in the enzyme (analogous to serine 195 in chymotrypsin) attacks the phosphate group to form a covalent intermediate. In a second step, this intermediate is hydrolyzed to produce the final products. In this case, two displacement reactions take place at the phosphorus atom. Consequently, the stereochemical configuration at the phosphorus atom would be inverted and then inverted again, and the overall configuration would be *retained*. In the second type of mechanism, analogous to that used by the aspartyl- and metalloproteases, an activated water molecule attacks the phosphorus atom directly. In this mechanism, a single displacement reaction takes place at the phosphorus atom. Hence, the stereochemical configuration at the phosphorus atom is *inverted* after cleavage. To determine which mechanism is correct, we examine the stereochemistry at the phosphorus atom after cleavage.

A difficulty is that the stereochemistry is not easily observed, because two of the groups bound to the phosphorus atom are simple oxygen atoms, identical with each other. This difficulty can be circumvented by replacing one oxygen atom with sulfur (producing a species called a phosphorothioate).

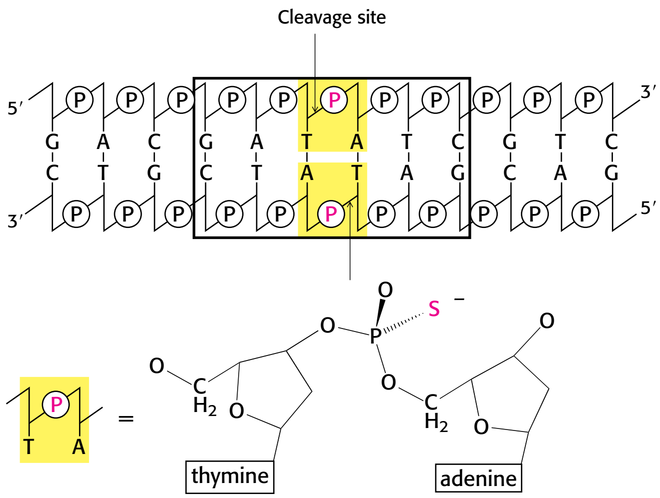


FIGURE 9.32 Labeling with phosphorothioates. Phosphorothioate groups, in which one of the nonbridging oxygen atoms is replaced by a sulfur atom, can be used to label specific sites in the DNA backbone to determine the overall stereochemical course of a displacement reaction. Here, a phosphorothioate is placed at sites that can be cleaved by EcoRV endonuclease.

Let us consider EcoRV endonuclease. This enzyme cleaves the phosphodiester bond between the T and the A at the center of the recognition sequence 5′-GATATC-3′. The first step is to synthesize an appropriate substrate for EcoRV containing phosphorothioates at the sites of cleavage (Figure 9.32). The reaction is then performed in water that has been greatly enriched in ^{18}O to allow the incoming oxygen atom to be marked. The location of the ^{18}O label with respect to the sulfur atom indicates whether the reaction proceeds with inversion or retention of stereochemistry. *This experiment revealed that the stereochemical configuration at the phosphorus atom was inverted only once with cleavage.* This result is consistent with a direct attack by water at the phosphorus atom and rules out the formation of any covalently bound intermediate (Figure 9.33).

Restriction enzymes require magnesium for catalytic activity

Many enzymes that act on phosphate-containing substrates require Mg^{2+} or some other similar divalent cation for activity. One or more Mg^{2+} (or similar) cations are essential to the function of restriction endonucleases. What are the functions of these metal ions?

Direct visualization of the complex between EcoRV endonuclease and cognate DNA molecules in the presence of Mg^{2+} by crystallization has not been possible, because the enzyme cleaves the substrate under these circumstances. Nonetheless, metal ion complexes can be visualized through several approaches. In one approach, crystals of EcoRV endonuclease are prepared bound to oligonucleotides that contain the enzyme's recognition sequence. These crystals are grown in the absence of magnesium to prevent cleavage; after their preparation, the crystals are soaked in solutions containing the metal. Alternatively, crystals have been grown with the use of a mutated form of the enzyme that is less active. Finally, Mg^{2+} can be replaced by

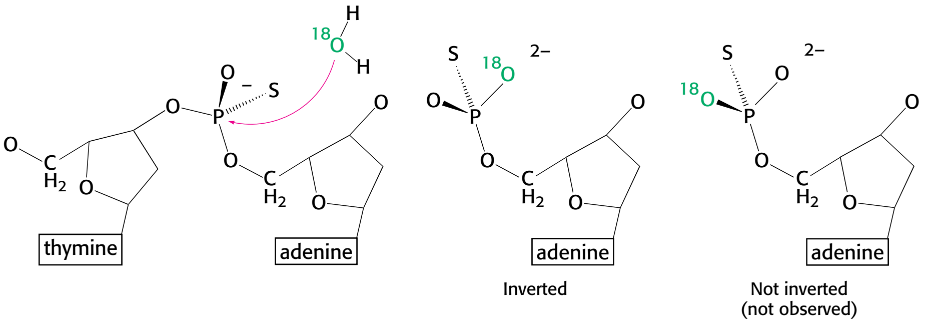


FIGURE 9.33 Stereochemistry of cleaved DNA. Cleavage of DNA by EcoRV endonuclease results in overall inversion of the stereochemical configuration at the phosphorus atom, as indicated by the stereochemistry of the phosphorus atom bound to one bridging oxygen atom, one ^{16}O, one ^{18}O, and one sulfur atom. Two possible products are shown, only one of which is observed, indicating direct attack of water at the phosphorous atom.

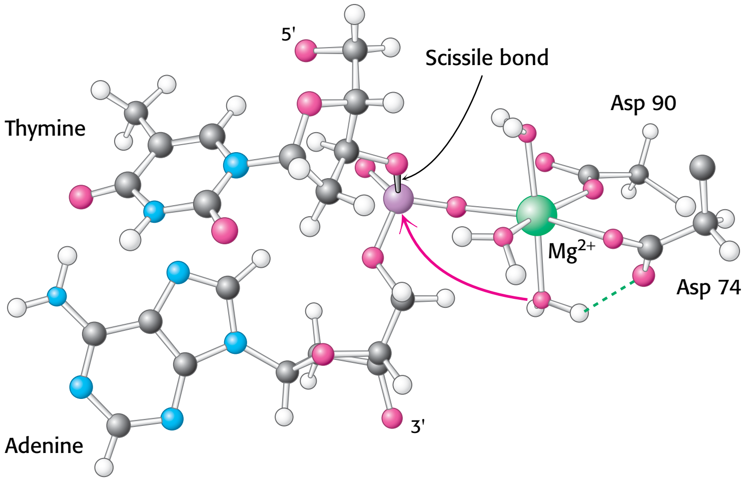


FIGURE 9.34 A magnesium ion-binding site in EcoRV endonuclease. The magnesium ion helps to activate a water molecule and positions it so that it can attack the phosphorus atom.

metal ions such as Ca^{2+} that bind but do not result in much catalytic activity. In all cases, no cleavage takes place, and so the locations of the metal ion-binding sites are readily determined.

As many as three metal ions have been found to be present per active site. The roles of these multiple metal ions are still under investigation. One ion-binding site is occupied in essentially all structures. This metal ion is coordinated to the protein through two aspartate residues and to one of the phosphate-group oxygen atoms near the site of cleavage. This metal ion binds the water molecule that attacks the phosphorus atom, helping to position and activate it in a manner similar to that for the Zn^{2+} ion of carbonic anhydrase (Figure 9.34).

The complete catalytic apparatus is assembled only within complexes of cognate DNA molecules, ensuring specificity

We now return to the question of specificity, the defining feature of restriction enzymes. The recognition sequences for most restriction endonucleases are *inverted repeats.* This arrangement gives the three-dimensional structure of the recognition site a *twofold rotational symmetry* (Figure 9.35).

The restriction enzymes display a corresponding symmetry: they are dimers whose two subunits are related by twofold rotational symmetry. The matching symmetry of the recognition sequence and the enzyme facilitates the recognition of cognate DNA by the enzyme. This similarity in structure has been confirmed by the determination of the structure of the complex between EcoRV endonuclease and DNA fragments containing its

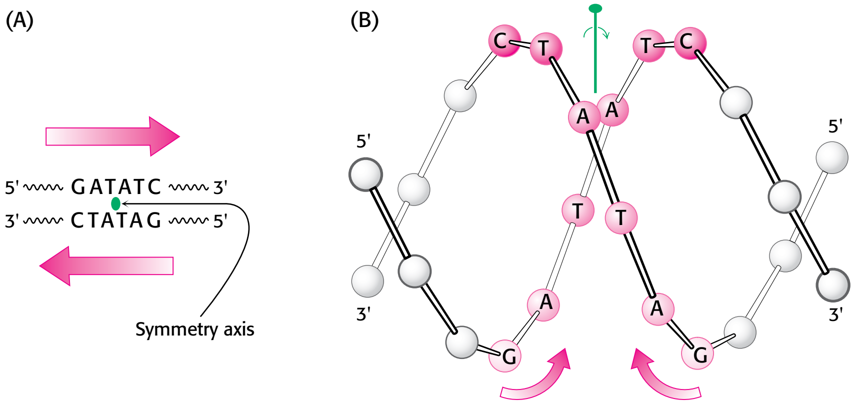


FIGURE 9.35 Structure of the recognition site of EcoRV endonuclease. (A) The sequence of the recognition site, which is symmetric around the axis of rotation designated in green. (B) The inverted repeat within the recognition sequence of EcoRV (and most other restriction endonucleases) endows the DNA site with twofold rotational symmetry.

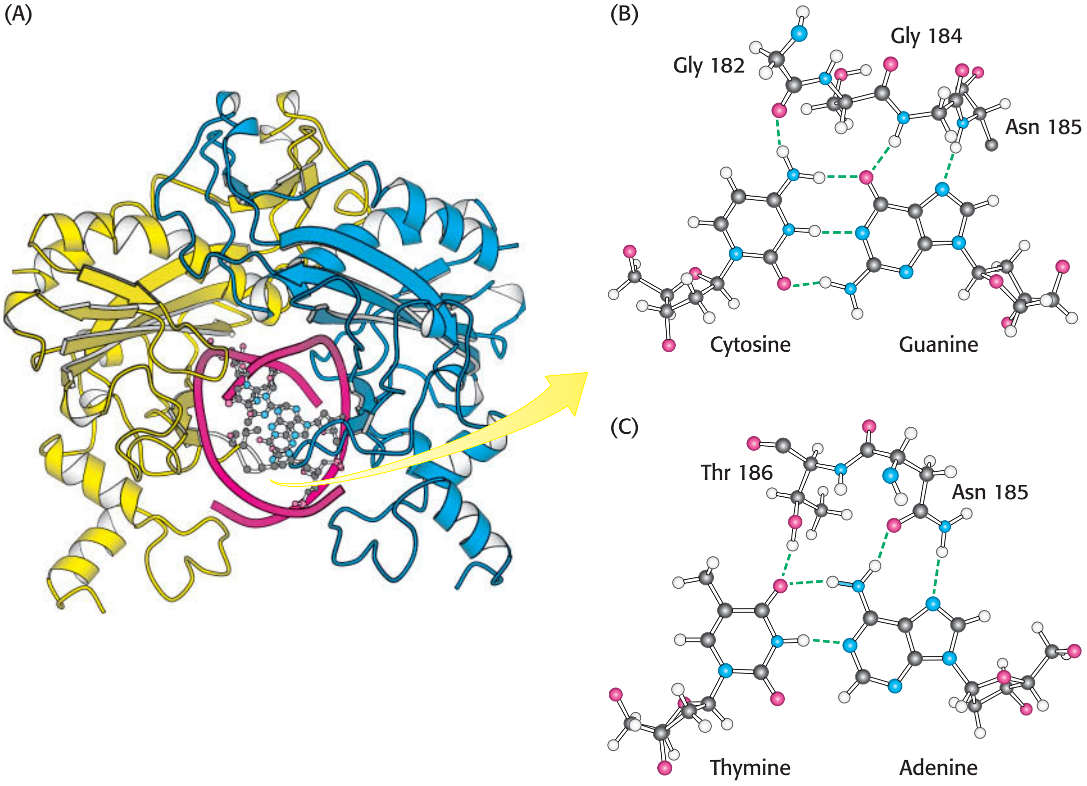


FIGURE 9.36 EcoRV embracing a cognate DNA molecule. (A) This view of the structure of EcoRV endonuclease bound to a cognate DNA fragment is down the helical axis of the DNA. The two protein subunits are in yellow and blue, and the DNA backbone is in red. *Notice* that the twofold axes of the enzyme dimer and the DNA are aligned. One of the DNA-binding loops (in green) of EcoRV endonuclease is shown interacting with the base pairs of its cognate DNA-binding site. Key amino acid residues are shown hydrogen bonding with (B) a CG base pair and (C) an AT base pair. [Drawn from 1RVB.pdb.]

recognition sequence (Figure 9.36). The enzyme surrounds the DNA in a tight embrace.

An enzyme's binding affinity for substrates often determines specificity. Surprisingly, however, binding studies performed in the absence of magnesium have demonstrated that the EcoRV endonuclease binds to all sequences, both cognate and noncognate, with approximately equal affinity. Why, then, does the enzyme cleave only cognate sequences? The answer lies in a unique set of interactions between the enzyme and a cognate DNA sequence.

Within the 5′-GATATC-3′ sequence, the G and A bases at the 5′ end of each strand and their Watson–Crick partners directly contact the enzyme by hydrogen bonding with residues that are located in two loops, one projecting from the surface of each enzyme subunit (Figure 9.36). The most striking feature of this complex is the *distortion of the DNA,* which is substantially kinked in the center (Figure 9.37). The central two TA base pairs in the recognition sequence play a key role in producing the kink. They do not make contact with the enzyme but appear to be required because of their ease of distortion. The 5′-TA-3′ sequence is known to be among the most easily deformed base pairs.

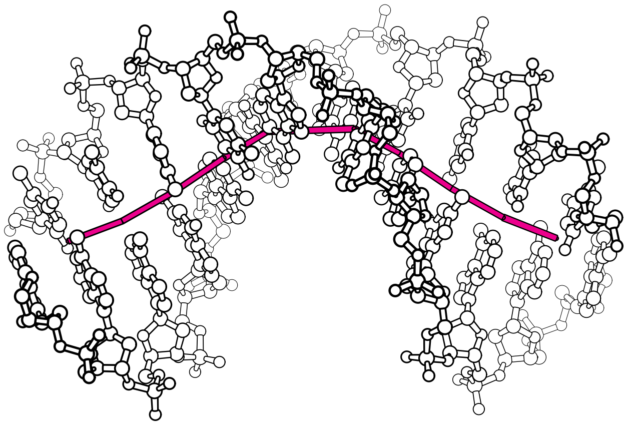


FIGURE 9.37 Distortion of the recognition site. The DNA is represented as a ball-and-stick model. The path of the DNA helical axis, shown in red, is substantially distorted on binding to the enzyme. For the B form of DNA, the axis is straight (not shown).

The structures of complexes formed with noncognate DNA fragments are strikingly different from those formed with cognate DNA:

the noncognate DNA conformation is not substantially distorted (Figure 9.38). *This lack of distortion has important consequences with regard to catalysis. No phosphate is positioned sufficiently close to the active-site aspartate residues to complete a magnesium ion-binding site* (Figure 9.34). Hence, the nonspecific complexes do not bind the magnesium ions and the complete catalytic apparatus is never assembled. The distortion of the substrate and the subsequent binding of the magnesium ion account for the catalytic specificity of more than a millionfold that is observed for EcoRV endonuclease. *Thus, enzyme specificity may be determined by the specificity of enzyme action rather than the specificity of substrate binding.*

We can now see the role of binding energy in this strategy for attaining catalytic specificity. The distorted DNA makes additional contacts with the enzyme, increasing the binding energy. However, the increase in binding energy is canceled by the energetic cost of distorting the DNA from its relaxed conformation (Figure 9.39). Thus, for EcoRV endonuclease, there is little difference in binding affinity for cognate and nonspecific DNA fragments. However, the distortion in the cognate complex dramatically affects catalysis by completing the magnesium ion-binding site. This example illustrates how enzymes can utilize available binding energy to deform substrates and poise them for chemical transformation. Interactions that take place within the distorted substrate complex stabilize the transition state leading to DNA hydrolysis.

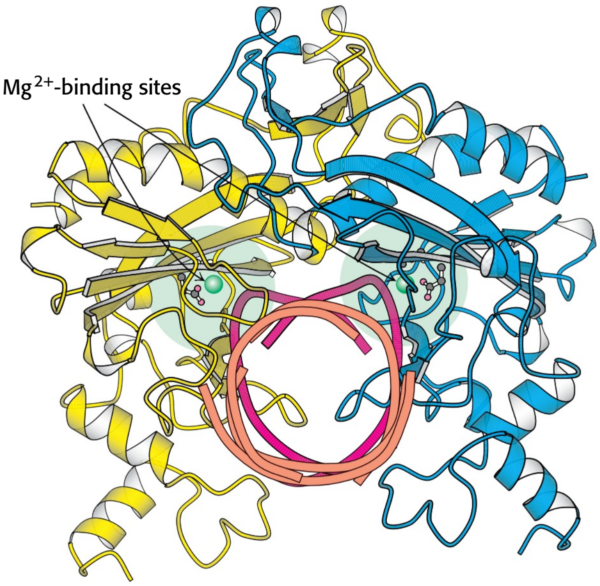


FIGURE 9.38 Nonspecific and cognate DNA within EcoRV endonuclease. A comparison of the positions of the nonspecific (orange) and the cognate DNA (red) within EcoRV. *Notice* that, in the nonspecific complex, the DNA backbone is too far from the enzyme to complete the magnesium ion-binding sites. [Drawn from 1RVB.pdb.]

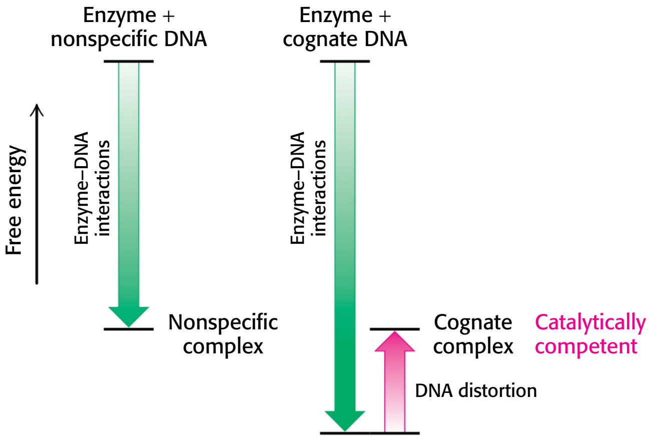


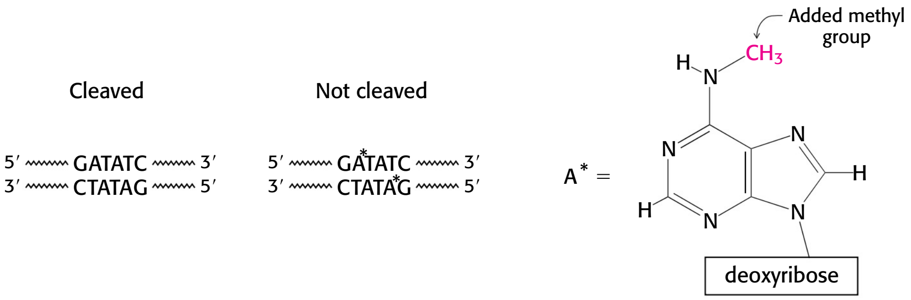
FIGURE 9.39 Greater binding energy of EcoRV endonuclease bound to cognate versus noncognate DNA. The additional interactions between EcoRV endonuclease and cognate DNA increase the binding energy, which can be used to drive DNA distortions necessary for forming a catalytically competent complex.

Host-cell DNA is protected by the addition of methyl groups to specific bases

How does a host cell harboring a restriction enzyme protect its own DNA? The host DNA is methylated on specific adenine bases within host recognition sequences by other enzymes called *methylases* (Figure 9.40). An endonuclease will not cleave DNA if its recognition sequence is methylated. For each restriction endonuclease, the host cell produces a corresponding methylase that marks the host DNA at the appropriate methylation site. These pairs of enzymes are referred to as *restriction-modification systems.*

The distortion in the DNA explains how methylation blocks catalysis and protects host-cell DNA. The host *E. coli* adds a methyl group to the amino group of the adenine nucleotide at the 5′ end of the recognition sequence. The presence of the methyl group blocks the formation of a hydrogen bond between the amino group and the side-chain

FIGURE 9.40 Protection by methylation. The recognition sequence for EcoRV endonuclease (left) and the sites of methylation (right) in DNA protected from the catalytic action of the enzyme.



carbonyl group of asparagine 185 (Figure 9.41). This asparagine residue is closely linked to the other amino acids that form specific contacts with the DNA. The absence of the hydrogen bond disrupts other interactions between the enzyme and the DNA substrate, and the distortion necessary for cleavage will not take place.

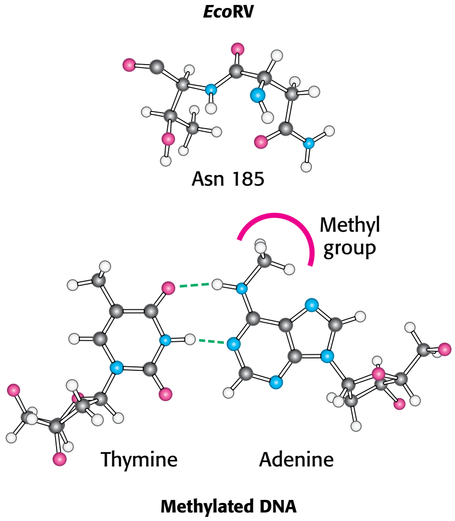


FIGURE 9.41 Methylation of adenine. The methylation of adenine blocks the formation of hydrogen bonds between EcoRV endonuclease and cognate DNA molecules and prevents their hydrolysis.

Type II restriction enzymes have a catalytic core in common and are probably related by horizontal gene transfer

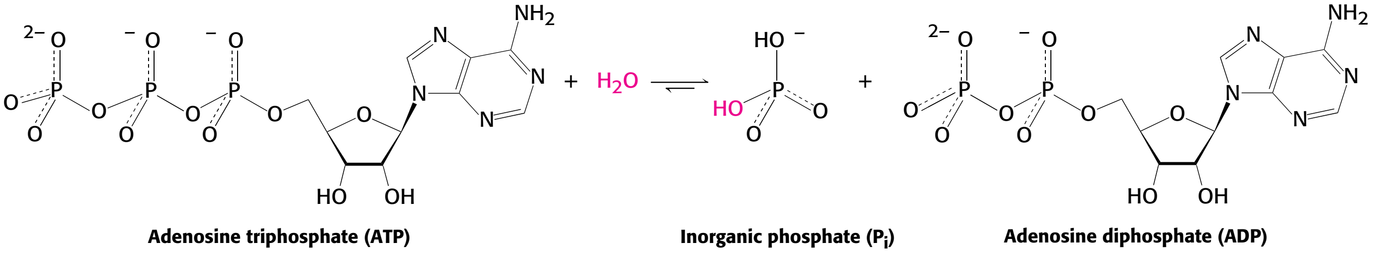
Type II restriction enzymes are prevalent in Archaea and Bacteria. What can we tell of the evolutionary history of these enzymes? Comparison of the amino acid sequences of a variety of type II restriction endonucleases did not reveal significant sequence similarity between most pairs of enzymes. However, a careful examination of three-dimensional structures, taking into account the location of the active sites, revealed the presence of a core structure conserved in the different enzymes.

These observations indicate that many type II restriction enzymes are indeed evolutionarily related. Analyses of the sequences in greater detail suggest that bacteria may have obtained genes encoding these enzymes from other species by *horizontal gene transfer,* the passing of pieces of DNA (such as plasmids) between species that provide a selective advantage in a particular environment. For example, EcoRI (from *E. coli*) and RsrI (from *Rhodobacter sphaeroides*) are 50% identical in sequence over 266 amino acids, clearly indicative of a close evolutionary relationship. However, these species of bacteria are not closely related. Thus, *these species appear to have obtained the gene for these restriction endonucleases from a common source more recently than the time of their evolutionary divergence.* Moreover, the codons used by the gene encoding EcoRI endonuclease to specify given amino acids are strikingly different from the codons used by most *E. coli* genes, which suggests that the gene did not originate in *E. coli.*

Horizontal gene transfer may be a common event. For example, genes that inactivate antibiotics are often transferred, leading to the transmission of antibiotic resistance from one species to another. For restriction-modification systems, protection against viral infections may have favored horizontal gene transfer.

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work

The final enzymes that we will consider are the myosins. These enzymes catalyze the hydrolysis of adenosine triphosphate (ATP) to form adenosine diphosphate (ADP) and inorganic phosphate (P_i) and use the energy associated with this thermodynamically favorable reaction to drive the motion of molecules within cells.


$2-$
$-$
$-$
O
P
NH_2
N
HO
OH
+
H_2O
HO
Adenosine triphosphate (ATP)
Inorganic phosphate (P_i)
Adenosine diphosphate (ADP)

For example, when we lift a book, the energy required comes from ATP hydrolysis catalyzed by myosin in our muscles. Myosins are found in all

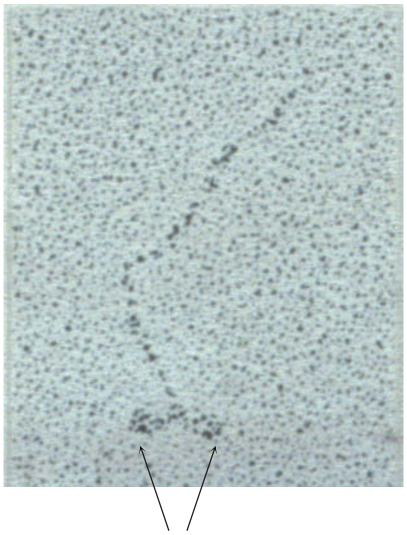


FIGURE 9.42 Elongated structure of muscle myosin. An electron micrograph showing myosin from mammalian muscle. This dimeric protein has an elongated structure with two globular ATPase domains per dimer. [Courtesy of Dr. Paula Flicker, Dr. Theo Walliman, and Dr. Peter Vibert.]

eukaryotes and the human genome encodes more than 40 different myosins. Myosins generally have elongated structures with globular domains that actually carry out ATP hydrolysis (Figure 9.42). In this chapter, we will focus on the globular ATPase domains, particularly the strategies that allow myosins to hydrolyze ATP in a controlled manner and to use the free energy associated with this reaction to promote substantial conformational changes within the myosin molecule. These conformational changes are amplified by other structures in the elongated myosin molecules to transport proteins or other cargo substantial distances within cells. In Chapter 35, we will examine the action of myosins and other molecular-motor proteins in much more detail.

As will be discussed in Chapter 15, ATP is used as the major currency of energy inside cells. Many enzymes use ATP hydrolysis to drive other reactions and processes. In almost all cases, an enzyme that hydrolyzed ATP without any such coupled processes would simply drain the energy reserves of a cell without benefit.

ATP hydrolysis proceeds by the attack of water on the gamma phosphoryl group

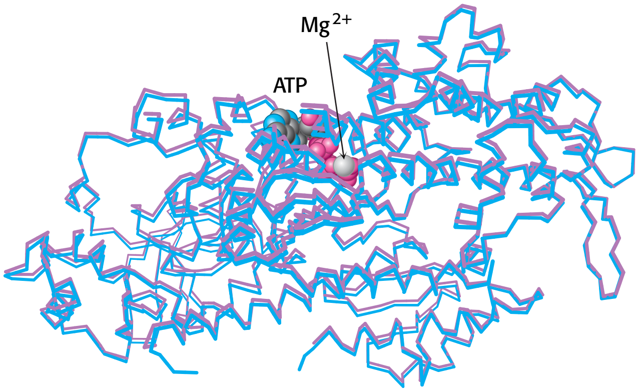
In our examination of the mechanism of restriction enzymes, we learned that an activated water molecule performs a nucleophilic attack on phosphorus to cleave the phosphodiester backbone of DNA. The cleavage of ATP by myosins follows an analogous mechanism. To understand the myosin mechanism in more detail, we must first examine the structure of the myosin *ATPase* domain.

The structures of the ATPase domains of several different myosins have been examined. One such domain, that from the soil-living amoeba *Dictyostelium discoideum,* an organism that has been extremely useful for studying cell movement and molecular-motor proteins, has been studied in great detail. The crystal structure of this protein fragment in the absence of nucleotides revealed a single globular domain comprising approximately 750 amino acids. A water-filled pocket is present toward the center of the structure, suggesting a possible nucleotide-binding site. Crystals of this protein were soaked in a solution containing ATP and the structure was examined again. Remarkably, this structure revealed intact ATP bound in the active site with very little change in the overall structure and without evidence of significant hydrolysis (Figure 9.43). The ATP is also bound to a Mg^{2+} ion.

Kinetic studies of myosins, as well as many other enzymes having ATP or other nucleoside triphosphates as a substrate, reveal that these enzymes are essentially inactive in the absence of divalent metal ions such as magnesium (Mg^{2+}) or manganese (Mn^{2+}) but acquire activity on the addition of these ions. In contrast with the enzymes discussed so far, the metal is not a component of the active site. Rather, nucleotides such as ATP bind these ions, and it is the metal ion–nucleotide complex that is the true substrate for the enzymes. The dissociation constant for the ATP–Mg^{2+} complex is approximately 0.1 mM, and thus, given that intracellular Mg^{2+} concentrations are typically in the millimolar range, essentially all nucleoside triphosphates are present as NTP–Mg^{2+} complexes. *Magnesium or manganese complexes of nucleoside triphosphates are the true substrates for essentially all NTP-dependent enzymes.*

The nucleophilic attack by a water molecule on the γ-phosphoryl group requires some mechanism to activate the water, such as a basic residue or a bound metal ion. Examination

FIGURE 9.43 Myosin–ATP complex structure. An overlay of the structures of the ATPase domain from *Dictyostelium discoideum* myosin with no ligands bound (blue) and the complex of this protein with ATP and magnesium bound (red). *Notice* that the two structures are extremely similar to one another. [Drawn from 1FMV.pdb and 1FMW.pdb].



of the myosin–ATP complex structure shows no basic residue in an appropriate position and reveals that the bound Mg^{2+} ion is too far away from the phosphoryl group to play this role. These observations suggest why this ATP complex is relatively stable; the enzyme is not in a conformation that is competent to catalyze the reaction. This observation suggests that the domain must undergo a conformational change to catalyze the ATP-hydrolysis reaction.

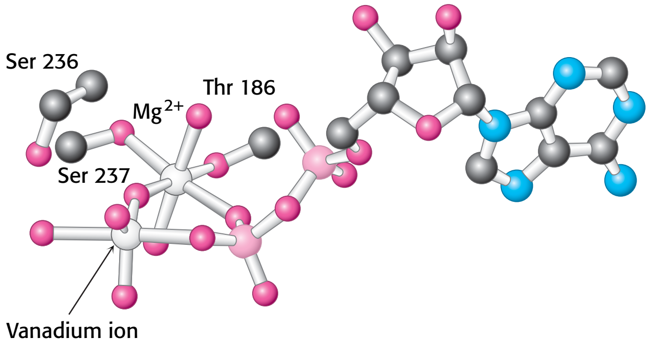
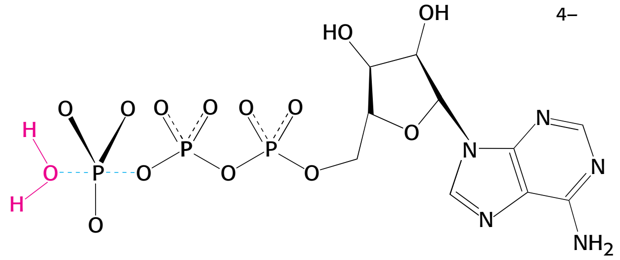


FIGURE 9.44 Myosin ATPase Transition-State Analog. The structure of the transition-state analog formed by treating the myosin ATPase domain with ADP and vanadate (VO_4^{3-}) in the presence of magnesium. *Notice* that the vanadium ion is coordinated to five oxygen atoms including one from ADP. The positions of two residues that bind magnesium as well as Ser 236, a residue that appears to play a direct role in catalysis, are shown. [Drawn from 1VOM.pdb]

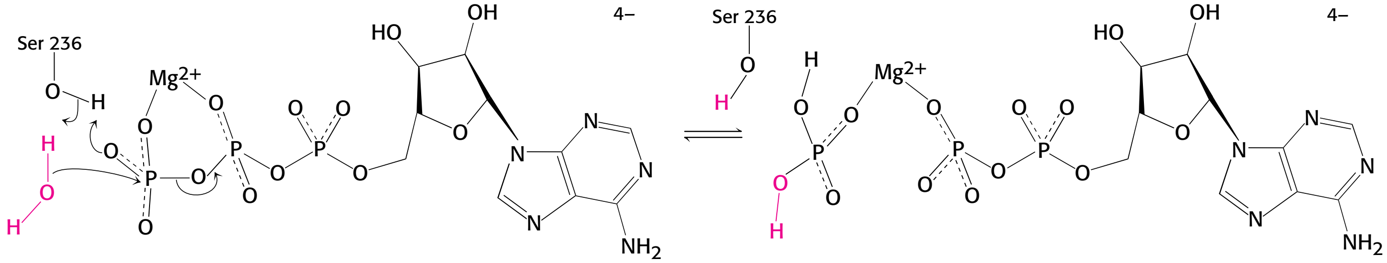
Formation of the transition state for ATP hydrolysis is associated with a substantial conformational change

The catalytically competent conformation of the myosin ATPase domain must bind and stabilize the transition state of the reaction. In analogy with restriction enzymes, we expect that ATP hydrolysis includes a pentacoordinate transition state.



Such pentacoordinate structures based on phosphorus are too unstable to be readily observed. However, transition-state analogs in which other atoms replace phosphorus are more stable. The transition metal vanadium, in particular, forms similar structures. The myosin ATPase domain can be crystallized in the presence of ADP and vanadate, VO_4^{3-}. The result is the formation of a complex that closely matches the expected transition-state structure (Figure 9.44). As expected, the vanadium atom is coordinated to five oxygen atoms, including one oxygen atom from ADP diametrically opposite an oxygen atom that is analogous to the attacking water molecule in the transition state. The Mg^{2+} ion is coordinated to one oxygen atom from the vanadate, one oxygen atom from the ADP, two hydroxyl groups from the enzyme, and two water molecules. In this position, this ion does not appear to play any direct role in activating the attacking water. However, an additional residue from the enzyme, Ser 236, is well positioned to play a role in catalysis (Figure 9.44). In the proposed mechanism of ATP hydrolysis based on this structure, the water molecule attacks the γ-phosphoryl group, with the hydroxyl group of Ser 236 facilitating the transfer of a proton from the attacking water to the hydroxyl group of Ser 236, which, in turn, is deprotonated by one of the oxygen atoms of the γ-phosphoryl group (Figure 9.45). *Thus, in effect, the ATP serves as a base to promote its own hydrolysis.*

FIGURE 9.45 Facilitating water attack. The water molecule attacking the γ-phosphoryl group of ATP is deprotonated by the hydroxyl group of Ser 236, which, in turn, is deprotonated by one of the oxygen atoms of the γ-phosphoryl group forming the $H_2PO_4^-$ product.



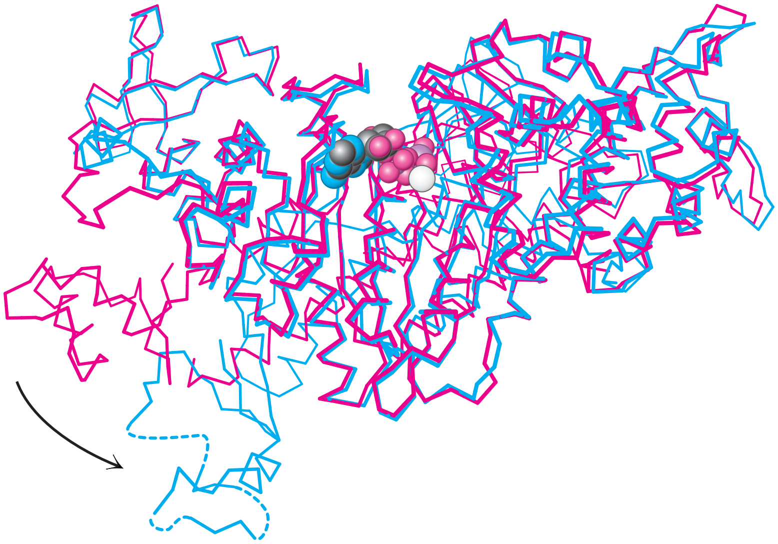


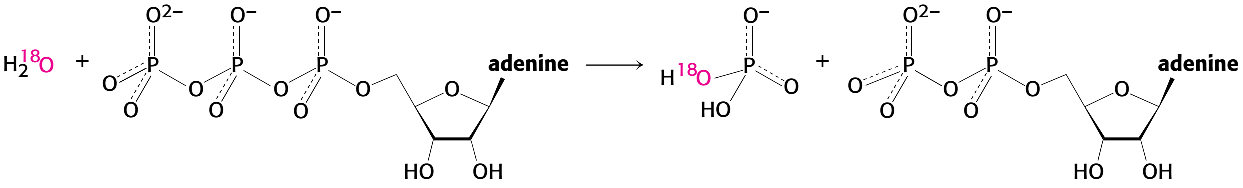
FIGURE 9.46 Myosin conformational changes. A comparison of the overall structures of the myosin ATPase domain with ATP bound (shown in red) and that with the transition-state analog ADP–vanadate (shown in blue). *Notice* the large conformational change of a region at the carboxyl-terminus of the domain, some parts of which move as much as 25 Å. [Drawn from 1FMW.pdb and 1VOM.pdb].

Comparison of the overall structures of the myosin ATPase domain complexed with ATP and with the ADP–vanadate reveals some remarkable differences. Relatively modest structural changes occur in and around the active site. In particular, a stretch of amino acids moves closer to the nucleotide by approximately 2 Å and interacts with the oxygen atom that corresponds to the attacking water molecule. These changes help facilitate the hydrolysis reaction by stabilizing the transition state. However, examination of the overall structure shows even more striking changes.

A region comprising approximately 60 amino acids at the carboxyl-terminus of the domain adopts a different configuration in the ADP–vanadate complex, displaced by as much as 25 Å from its position in the ATP complex (Figure 9.46). This displacement tremendously amplifies the relatively subtle changes that take place in the active site. The effect of this motion is amplified even more as this carboxyl-terminal domain is connected to other structures within the elongated structures typical of myosin molecules (Figure 9.42). Thus, the conformation that is capable of promoting the ATP hydrolysis reaction is itself substantially different from other conformational changes that take place in the course of the catalytic cycle.

The altered conformation of myosin persists for a substantial period of time

Myosins are slow enzymes, typically turning over approximately once per second. What steps limit the rate of turnover? In an experiment that was particularly revealing, the hydrolysis of ATP was catalyzed by the myosin ATPase domain from mammalian muscle. The reaction took place in water labeled with ^{18}O to track the incorporation of solvent oxygen into the reaction products. The fraction of oxygen in the phosphate product was analyzed. In the simplest case, the phosphate would be expected to contain one oxygen atom derived from water and three initially present in the terminal phosphoryl group of ATP.



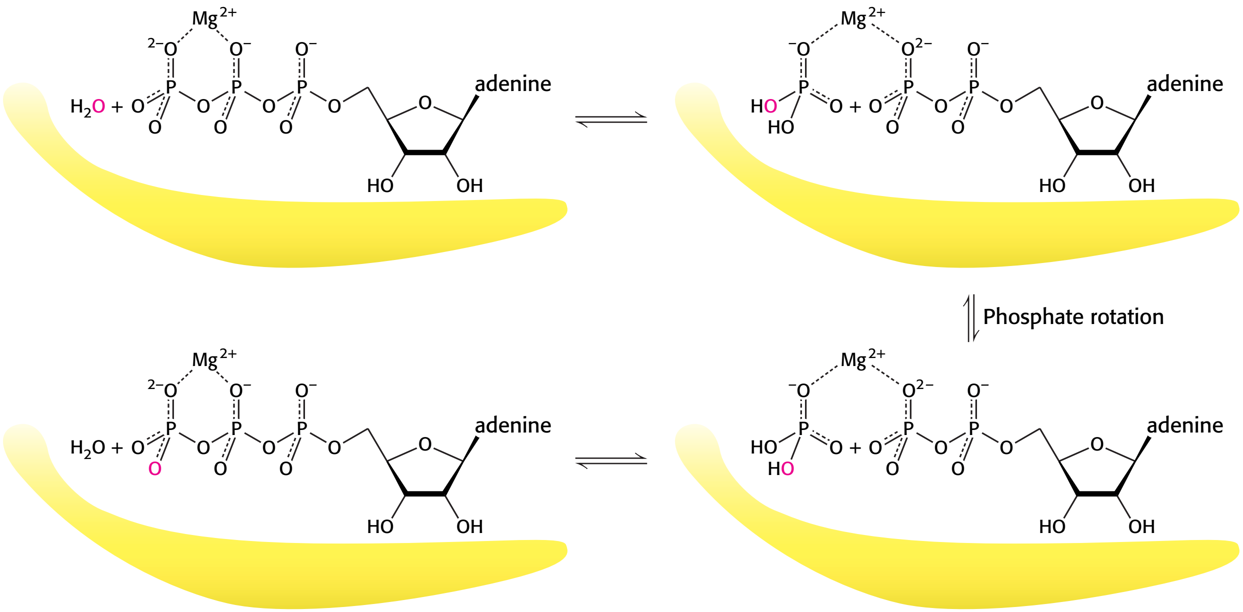


FIGURE 9.47 Reversible hydrolysis of ATP within the myosin active site. For myosin, more than one atom of oxygen from water is incorporated in inorganic phosphate. The oxygen atoms are incorporated in cycles of hydrolysis of ATP to ADP and inorganic phosphate, phosphate rotation within the active site, and reformation of ATP now containing oxygen from water.

Instead, between two and three of the oxygen atoms in the phosphate were found, on average, to be derived from water. These observations indicate that the ATP hydrolysis reaction within the enzyme active site is reversible. Each molecule of ATP is cleaved to ADP and P_i and then re-formed from these products several times before the products are released from the enzyme (Figure 9.47). At first glance, this observation is startling because ATP hydrolysis is a very favorable reaction with an equilibrium constant of approximately 140,000. However, this equilibrium constant applies to the molecules free in solution, not within the active site of an enzyme. Indeed, more-extensive analysis suggests that this equilibrium constant on the enzyme is approximately 10, indicative of a general strategy used by enzymes. Enzymes catalyze reactions by stabilizing the transition state. The structure of this transition state is intermediate between the enzyme-bound reactants and the enzyme-bound products. Many of the interactions that stabilize the transition state will help equalize the stabilities of the reactants and the products. Thus, *the equilibrium constant between enzyme-bound reactants and products is often close to 1, regardless of the equilibrium constant for the reactants and products free in solution.*

These observations reveal that the hydrolysis of ATP to ADP and P_i is not the rate-limiting step for the reaction catalyzed by myosin. Instead, the release of the products, particularly P_i, from the enzyme is rate limiting. The fact that a conformation of myosin with ATP hydrolyzed but still bound to the enzyme persists for a significant period of time is critical for coupling conformational changes that take place in the course of the reaction to other processes.

Scientists can watch single molecules of myosin move

Myosin molecules function to use the free energy of hydrolysis of ATP to drive macroscopic motion. Myosin molecules move along a filamentous protein termed actin, as we will discuss in more detail in Chapter 35. Using a variety of physical methods, scientists have been able to watch *single myosin in action*. For example, a myosin family member termed myosin V can be labelled with fluorescent tags so that it can be localized when fixed on a surface with a precision of less than 15 Å. When this myosin is placed on a surface coated with actin filaments, each molecule remains in a fixed position. However, when ATP is added, each molecule moves along the surface.

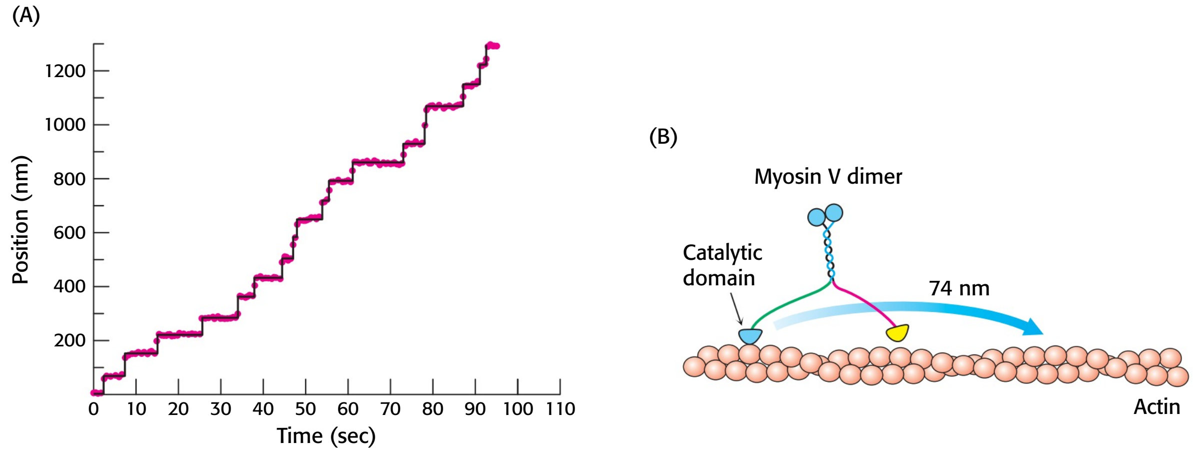


FIGURE 9.48 Single molecule motion. (A) A trace of the position of a single dimeric myosin V molecule as it moves across a surface coated with actin filaments. (B) A model of how the dimeric molecule moves in discrete steps with an average size of 74 ± 5 nm. [Data from A. Yildiz, et al., *Science* 300(5628):2061–2065, 2003.]

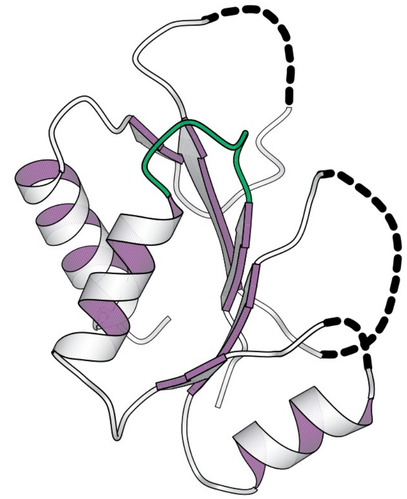
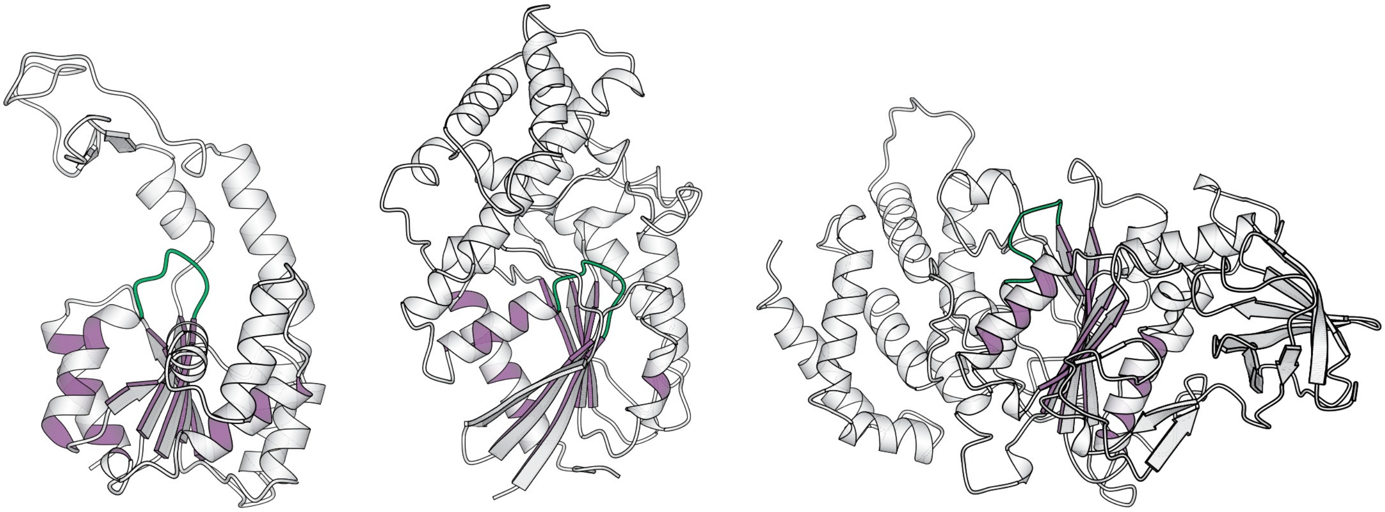


FIGURE 9.49 The core domain of NMP kinases. *Notice* the P-loop shown in green. The dashed lines represent the remainder of the protein structure. [Drawn from 1GKY.pdb.]

Tracking individual molecules reveals that each moves in steps of approximately 74 nm as shown in Figure 9.48. The observation of steps of a fixed size as well as the determination of this step size helps reveal details of the mechanism of action of these tiny molecular motors.

Myosins are a family of enzymes containing P-loop structures

X-ray crystallography has yielded the three-dimensional structures of a number of different enzymes that share key structural characteristics and, almost certainly, an evolutionary history with myosin. In particular, a conserved NTP-binding core domain is present. This domain consists of a central β sheet, surrounded on both sides by α helices (Figure 9.49). A characteristic feature of this domain is a loop between the first β strand and the first helix. This loop typically has several glycine residues that are often conserved between more closely related members of this large and diverse



Adenylate kinase **α subunit of transducin** **β subunit of ATP synthase**

FIGURE 9.50 Three proteins containing P-loop NTPase domains. *Notice* the conserved domains shown with the inner surfaces of the ribbons in purple and the P-loops in green. [Drawn from 4AKE.pdb; 1TND.pdb; 1BMF.pdb.]

family. The loop is often referred to as the *P-loop* because it interacts with phosphoryl groups on the bound nucleotide. P-loop NTPase domains are present in a remarkably wide array of proteins, many of which participate in essential biochemical processes. Examples include ATP synthase, the key enzyme responsible for ATP generation; signal-transduction proteins such as G proteins; proteins essential for translating mRNA into proteins, such as elongation factor Tu; and DNA- and RNA-unwinding helicases. The wide utility of P-loop NTPase domains is perhaps best explained by their ability to undergo substantial conformational changes on nucleoside triphosphate binding and hydrolysis. We shall encounter these domains throughout the book and shall observe how they function as springs, motors, and clocks. To allow easy recognition of these domains in the book, they will be depicted with the inner surfaces of the ribbons in a ribbon diagram shown in purple and the P-loop shown in green (Figure 9.50).

SUMMARY

Enzymes adopt conformations that are structurally and chemically complementary to the transition states of the reactions that they catalyze. Sets of interacting amino acid residues make up sites with the special structural and chemical properties necessary to stabilize the transition state. Enzymes use five basic strategies to form and stabilize the transition state. The first (1) involves the use of binding energy to promote both specificity and catalysis. The others are (2) covalent catalysis, (3) general acid–base catalysis, (4) catalysis by approximation, and (5) metal ion catalysis. The four classes of enzymes examined in this chapter catalyze the addition of water to their substrates but have different requirements for catalytic speed, specificity, and coupling to other processes.

9.1 Proteases Facilitate a Fundamentally Difficult Reaction

The cleavage of peptide bonds by chymotrypsin is initiated by the attack by a serine residue on the peptide carbonyl group. The attacking hydroxyl group is activated by interaction with the imidazole group of a histidine residue, which is, in turn, linked to an aspartate residue. This Ser-His-Asp catalytic triad generates a powerful nucleophile. The product of this initial reaction is a covalent intermediate formed by the enzyme and an acyl group derived from the bound substrate. The hydrolysis of this acyl-enzyme intermediate completes the cleavage process. The tetrahedral intermediates for these reactions have a negative charge on the peptide carbonyl oxygen atom. This negative charge is stabilized by interactions with peptide NH groups in a region on the enzyme termed the oxyanion hole.

Other proteases employ the same catalytic strategy. Some of these proteases, such as trypsin and elastase, are homologs of chymotrypsin. Other proteases, such as subtilisin, contain a very similar catalytic triad that has arisen by convergent evolution. Active-site structures that differ from the catalytic triad are present in a number of other classes of proteases. These classes employ a range of catalytic strategies but, in each case, a nucleophile is generated that is sufficiently powerful to attack the peptide carbonyl group. In some enzymes, the nucleophile is derived from a side chain whereas, in others, an activated water molecule attacks the peptide carbonyl directly.

9.2 Carbonic Anhydrases Make a Fast Reaction Faster

Carbonic anhydrases catalyze the reaction of water with carbon dioxide to generate carbonic acid. The catalysis can be extremely fast: some carbonic anhydrases hydrate carbon dioxide at rates as high as 1 million times per second. A tightly bound zinc ion is a crucial component of the active sites of these enzymes. Each zinc ion binds a water molecule and promotes its deprotonation to generate a hydroxide ion at neutral pH. This hydroxide ion attacks carbon dioxide to form bicarbonate ion, HCO_3^-. Because of the physiological roles of carbon dioxide and bicarbonate ions, speed is of the essence for this enzyme. To overcome limitations imposed by the rate of proton transfer from the zinc-bound water molecule, the most-active carbonic anhydrases have evolved a proton shuttle to transfer protons to a buffer.

9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions

A high level of substrate specificity is often the key to biological function. Restriction endonucleases that cleave DNA at specific recognition sequences discriminate between molecules that contain these recognition sequences and those that do not. Within the enzyme–substrate complex, the DNA substrate is distorted in a manner that generates a magnesium ion-binding site between the enzyme and DNA. The magnesium ion binds and activates a water molecule, which attacks the phosphodiester backbone.

Some enzymes discriminate between potential substrates by binding them with different affinities. Others may bind many potential substrates but promote chemical reactions efficiently only on specific molecules. Restriction endonucleases such as EcoRV endonuclease employ the latter mechanism. Only molecules containing the proper recognition sequence are distorted in a manner that allows magnesium ion binding and, hence, catalysis. Restriction enzymes are prevented from acting on the DNA of a host cell by the methylation of key sites within its recognition sequences. The added methyl groups block specific interactions between the enzymes and the DNA such that the distortion necessary for cleavage does not take place.

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work

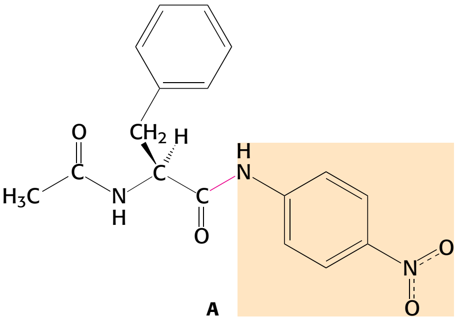
Finally, myosins catalyze the hydrolysis of adenosine triphosphate (ATP) to form adenosine diphosphate (ADP) and inorganic phosphate (P_i). The conformations of myosin ATPase domains free of bound nucleotides and with bound ATP are quite similar. Through the use of ADP and vanadate (VO_4^{3-}), an excellent mimic of the transition state for ATP hydrolysis bound to the myosin ATPase domain can be produced. The structure of this complex reveals that dramatic conformational changes take place on formation of this species from the ATP complex. These conformational changes are used to drive substantial motions in molecular motors. The rate of ATP hydrolysis by myosin is relatively low and is limited by the rate of product release from the enzyme. The hydrolysis of ATP to ADP and P_i within the enzyme is reversible with an equilibrium constant of approximately 10, compared with an equilibrium constant of 140,000 for these species free in solution. Myosins are examples of P-loop NTPase enzymes, a large collection of protein families that play key roles in a range of biological processes by virtue of the conformational changes that they undergo with various nucleotides bound.

KEY TERMS

binding energy (p. 252)
induced fit (p. 252)
covalent catalysis (p. 252)
general acid-base catalysis (p. 252)
catalysis by approximation (p. 252)
metal ion catalysis (p. 252)
chemical modification reaction (p. 254)
chromogenic substrate (p. 254)
catalytic triad (p. 256)
oxyanion hole (p. 257)
protease inhibitor (p. 263)
proton shuttle (p. 267)
recognition sequence (p. 269)
in-line displacement (p. 270)
methylases (p. 274)
restriction-modification system (p. 274)
horizontal gene transfer (p. 275)
ATPase (p. 276)
P-loop (p. 281)

PROBLEMS

1. *No burst.* Examination of the cleavage of the amide substrate, A, by chymotrypsin with the use of stopped-flow kinetic methods reveals no burst. The reaction is monitored by noting the color produced by the release of the amino part of the substrate (highlighted in orange). Why is no burst observed?



A

2. *Contributing to your own demise.* Consider the subtilisin substrates A and B.

Phe-Ala-Gln-Phe-X	Phe-Ala-His-Phe-X
A	**B**

These substrates are cleaved (between Phe and X) by native subtilisin at essentially the same rate. However, the His 64-to-Ala mutant of subtilisin cleaves substrate B more than 1000-fold as rapidly as it cleaves substrate A. Propose an explanation.

3. $1 + 1 \neq 2$. Consider the following argument. In subtilisin, mutation of Ser 221 to Ala results in a 10^6-fold decrease in activity. Mutation of His 64 to Ala results in a similar 10^6-fold decrease. Therefore, simultaneous mutation of Ser 221 to Ala and His 64 to Ala should result in a $10^6 \times 10^6 = 10^{12}$-fold reduction in activity. Is this reduction correct? Why or why not?

4. *Adding a charge.* In chymotrypsin, a mutant was constructed with Ser 189, which is in the bottom of the substrate-specificity pocket, changed to Asp. What effect would you predict for this Ser 189→Asp 189 mutation?

5. *Conditional results.* In carbonic anhydrase II, mutation of the proton-shuttle residue His 64 to Ala was expected to result in a decrease in the maximal catalytic rate. However, in buffers such as imidazole with relatively small molecular components, no rate reduction was observed. In buffers with larger molecular components, significant rate reductions were observed. Propose an explanation.

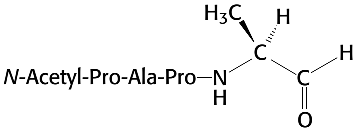
6. *How many sites?* A researcher has isolated a restriction endonuclease that cleaves at only one particular 10-base-pair site. Would this enzyme be useful in protecting cells from viral infections, given that a typical viral genome is 50,000 base pairs long? Explain.

7. *Is faster better?* Restriction endonucleases are, in general, quite slow enzymes with typical turnover numbers of $1\ s^{-1}$. Suppose that endonucleases were faster, with turnover numbers similar to those for carbonic anhydrase ($10^6\ s^{-1}$), such that they act faster than do methylases. Would this increased rate be beneficial to host cells, assuming that the fast enzymes have similar levels of specificity?

8. *Adopting a new gene.* Suppose that one species of bacteria obtained one gene encoding a restriction endonuclease by horizontal gene transfer. Would you expect this acquisition to be beneficial?

9. *Chelation therapy.* Treatment of carbonic anhydrase with high concentrations of the metal chelator EDTA (ethylenediaminetetraacetic acid) results in the loss of enzyme activity. Propose an explanation.

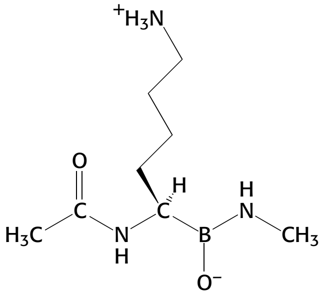
10. *An aldehyde inhibitor.* Elastase is specifically inhibited by an aldehyde derivative of one of its substrates:



(a) Which residue in the active site of elastase is most likely to form a covalent bond with this aldehyde?

(b) What type of covalent link would be formed?

11. *Identify the enzyme*. Consider the structure of molecule A. Which enzyme discussed in the chapter do you think molecule A will most effectively inhibit?



Molecule A

12. *Acid test*. At pH 7.0, carbonic anhydrase exhibits a k_{cat} of 600,000 s^{-1}. Estimate the value expected for k_{cat} at pH 6.0.

13. *Restriction*. To terminate a reaction in which a restriction enzyme cleaves DNA, researchers often add high concentrations of the metal chelator EDTA (ethylenediaminetetraacetic acid). Why does the addition of EDTA terminate the reaction?

14. *Labeling strategy*. ATP is added to the myosin ATPase domain in water labeled with ^{18}O. After 50% of the ATP has been hydrolyzed, the remaining ATP is isolated and found to contain ^{18}O. Explain.

15. *Viva la resistance*. Many patients become resistant to HIV protease inhibitors with the passage of time owing to mutations in the HIV gene that encodes the protease. Mutations are not found in the aspartate residue that interacts with the drugs. Why not?

16. *More than one way to skin k_{cat}*. Serine 236 in *Dictyostelium discoideum* myosin has been mutated to alanine. The mutated protein showed modestly reduced ATPase activity. Analysis of the crystal structure of the mutated protein revealed that a water molecule occupied the position of the hydroxyl group of the serine residue in the wild-type protein. Propose a mechanism for the ATPase activity of the mutated enzyme.

17. *A power struggle*. The catalytic power of an enzyme can be defined as the ratio of the rate of the enzyme catalyzed reaction to that for the uncatalyzed reaction. Using the information in Figure 9.15 for subtilisin and in Figure 9.22 for carbonic anhydrase, calculate the catalytic powers for these two enzymes.

18. *Wounded but not dead*. How much activity (in terms of relative k_{cat} values) does the version of subtilisin with all three residues in the catalytic triad mutated have compared to uncatalyzed reaction? Propose an explanation.

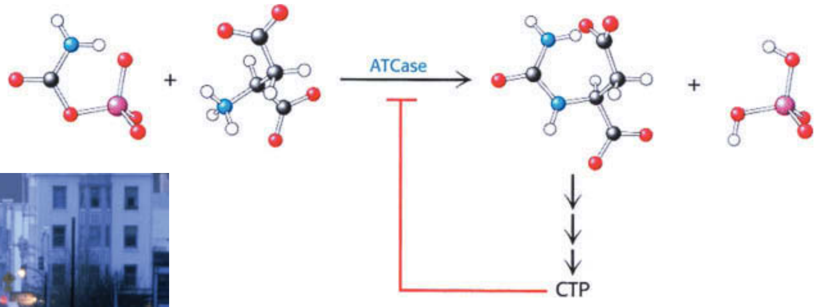
Mechanism Problem

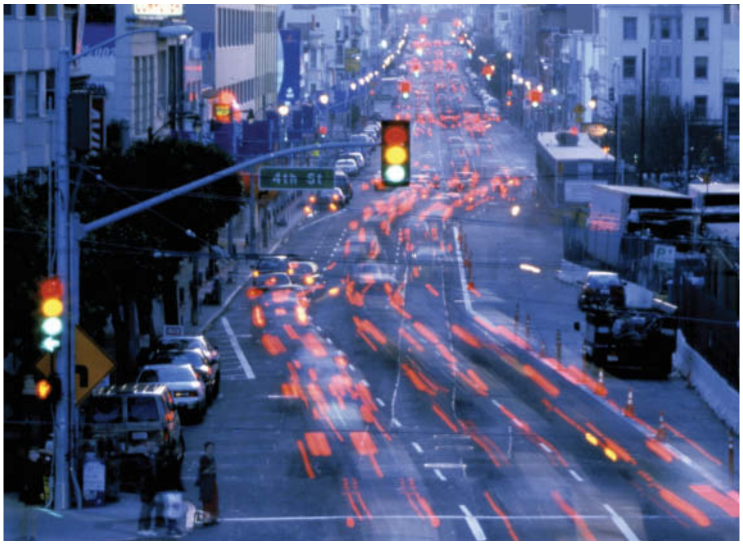
19. *Complete the mechanism*. On the basis of the information provided in Figure 9.17, complete the mechanisms for peptide-bond cleavage by (a) a cysteine protease, (b) an aspartyl protease, and (c) a metalloprotease.

CHAPTER

10

Regulatory Strategies





Like motor traffic, metabolic pathways flow more efficiently when regulated by signals. Cytidine triphosphate (CTP), the final product of a multistep pathway, controls flux through the pathway by inhibiting the committed step catalyzed by aspartate transcarbamoylase (ATCase). [(Left) Michael Winokur Photography/Getty Images.]

OUTLINE

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity

10.4 Many Enzymes Are Activated by Specific Proteolytic Cleavage

The activity of enzymes must often be regulated so that they function at the proper time and place. This regulation is essential for coordination of the vast array of biochemical processes taking place at any instant in an organism. Enzymatic activity is regulated in five principal ways:

1. *Allosteric Control.* Allosteric proteins contain distinct regulatory sites and multiple functional sites. The binding of small signal molecules at regulatory sites controls the activity of these proteins. Moreover, allosteric proteins show the property of *cooperativity:* activity at one functional site affects the activity at others. Proteins displaying allosteric control are thus information transducers: their activity can be modified in response to signal molecules or to information shared among active sites. This chapter examines one of the best-understood allosteric proteins: the enzyme *aspartate transcarbamoylase* (ATCase). Catalysis by aspartate transcarbamoylase of the first step in pyrimidine biosynthesis is inhibited by cytidine triphosphate, the final product of that biosynthesis, in an example of *feedback inhibition.* We have already examined an allosteric protein—hemoglobin, the oxygen transport protein in the blood (Chapter 7).

2. *Multiple Forms of Enzymes.* Isozymes, or isoenzymes, provide an avenue for varying regulation of the same reaction to meet the specific physiological needs in the particular tissue at a particular time. Isozymes are homologous enzymes within a single organism that catalyze the same reaction but differ

slightly in structure and more obviously in K_M and V_{max} values as well as in regulatory properties. Often, isozymes are expressed in a distinct tissue or organelle or at a distinct stage of development.

3. *Reversible Covalent Modification.* The catalytic properties of many enzymes are markedly altered by the covalent attachment of a modifying group, commonly a phosphoryl group. ATP serves as the phosphoryl donor in these reactions, which are catalyzed by *protein kinases.* The removal of phosphoryl groups by hydrolysis is catalyzed by *protein phosphatases.* This chapter considers the structure, specificity, and control of *protein kinase A* (PKA), a ubiquitous eukaryotic enzyme that regulates diverse target proteins.

4. *Proteolytic Activation.* The enzymes controlled by some of these regulatory mechanisms cycle between active and inactive states. A different regulatory strategy is used to *irreversibly* convert an inactive enzyme into an active one. Many enzymes are activated by the hydrolysis of a few peptide bonds or even one such bond in inactive precursors called *zymogens* or *proenzymes.* This regulatory mechanism generates digestive enzymes such as chymotrypsin, trypsin, and pepsin. Blood clotting is due to a remarkable cascade of zymogen activations. Active digestive and clotting enzymes are switched off by the irreversible binding of specific inhibitory proteins that are irresistible lures to their molecular prey.

5. *Controlling the Amount of Enzyme Present*. Enzyme activity can also be regulated by adjusting the amount of enzyme present. This important form of regulation usually takes place at the level of transcription. We will consider the control of gene transcription in Chapters 29, 30, and 31.

To begin, we will consider the principles of allostery by examining the enzyme aspartate transcarbamoylase.

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway

Aspartate transcarbamoylase catalyzes the first step in the biosynthesis of pyrimidines: the condensation of aspartate and carbamoyl phosphate to form *N*-carbamoylaspartate and orthophosphate (Figure 10.1). This reaction is the committed step in the pathway, which consists of 10 reactions, that will ultimately yield the pyrimidine nucleotides uridine triphosphate (UTP)

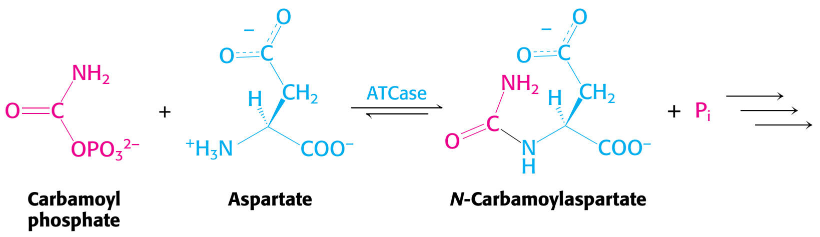
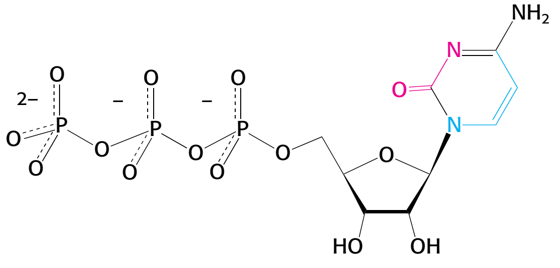


FIGURE 10.1 ATCase reaction. Aspartate transcarbamoylase catalyzes the committed step, the condensation of aspartate and carbamoyl phosphate to form *N*-carbamoylaspartate, in pyrimidine synthesis.



Cytidine triphosphate (CTP)

and cytidine triphosphate (CTP). How is this enzyme regulated to generate precisely the amount of pyrimidines needed by the cell?

ATCase is inhibited by CTP, the final product of the ATCase-initiated pathway. The rate of the reaction catalyzed by ATCase is fast at low concentrations of CTP but slows as CTP concentration increases (Figure 10.2). Thus, the pathway continues to make new pyrimidines until sufficient quantities of CTP have accumulated. The inhibition of ATCase by CTP is an example of *feedback inhibition,* the inhibition of an enzyme by the end product of the pathway. Feedback inhibition by CTP ensures that *N*-carbamoylaspartate and subsequent intermediates in the pathway are not needlessly formed when pyrimidines are abundant.

The inhibitory ability of CTP is remarkable because *CTP is structurally quite different from the substrates of the reaction* (Figure 10.1). Thus CTP must bind to a site distinct from the active site at which substrate binds. Such sites are called *allosteric* or *regulatory sites.* CTP is an example of an *allosteric inhibitor.* In ATCase (but not all allosterically regulated enzymes), the catalytic sites and the regulatory sites are on separate polypeptide chains.

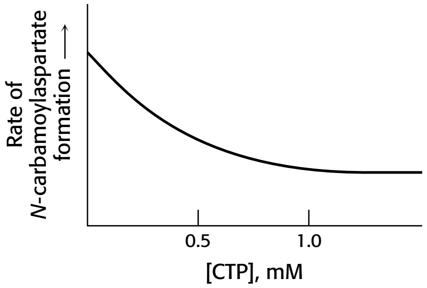


FIGURE 10.2 CTP inhibits ATCase. Cytidine triphosphate, an end product of the pyrimidine-synthesis pathway, inhibits aspartate transcarbamoylase despite having little structural similarity to reactants or products.

Allosterically regulated enzymes do not follow Michaelis–Menten kinetics

Allosteric enzymes are distinguished by their response to changes in substrate concentration in addition to their susceptibility to regulation by other molecules. Let us examine the rate of product formation as a function of substrate concentration for ATCase (Figure 10.3). The curve differs from that expected for an enzyme that follows Michaelis–Menten kinetics. The observed curve is referred to as sigmoidal because it resembles the letter "S." The vast majority of allosteric enzymes display sigmoidal kinetics. Recall from the discussion of hemoglobin that sigmoidal curves result from cooperation between subunits: the binding of substrate to one active site in a molecule increases the likelihood that substrate will bind to other active sites. To understand the basis of sigmoidal enzyme kinetics and inhibition by CTP, we need to examine the structure of ATCase.

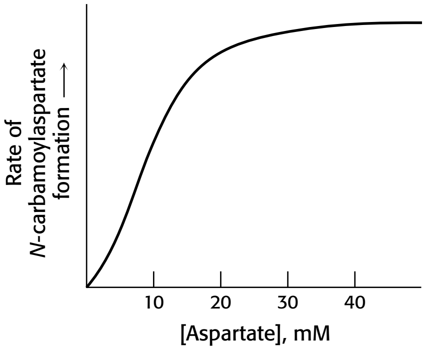


FIGURE 10.3 ATCase displays sigmoidal kinetics. A plot of product formation as a function of substrate concentration produces a sigmoidal curve because the binding of substrate to one active site increases the activity at the other active sites. Thus, the enzyme shows cooperativity.

ATCase consists of separable catalytic and regulatory subunits

What is the evidence that ATCase has distinct regulatory and catalytic sites? ATCase can be literally separated into regulatory (r) and catalytic (c) subunits by treatment with a mercurial compound such as *p*-hydroxymercuribenzoate, which reacts with sulfhydryl groups (Figure 10.4). Ultracentrifugation (p. 74) following treatment with mercurials revealed that ATCase is composed of two kinds of subunits (Figure 10.5). The subunits can be readily separated by ion-exchange chromatography because they differ markedly in charge (p. 67) or by centrifugation in a sucrose density gradient because they differ in size (p. 67). These size differences are manifested in the sedimentation coefficients: that of the native enzyme is 11.6S, whereas those of the dissociated subunits are 2.8S and 5.8S. The attached *p*-mercuribenzoate groups can be removed from the separated subunits by adding an excess of mercaptoethanol, providing isolated subunits for study.

The larger subunit is the *catalytic subunit.* This subunit has catalytic activity but displays the hyperbolic kinetics of Michaelis-Menten enzymes rather than sigmoidal kinetics. Furthermore, the isolated catalytic subunit is unresponsive to CTP. The isolated smaller subunit can bind CTP, but has no catalytic activity. Hence, that subunit is the *regulatory subunit.* The catalytic subunit (c_3) consists of three chains (34 kDa each), and the regulatory subunit (r_2) consists of two chains (17 kDa each). The catalytic and regulatory subunits combine

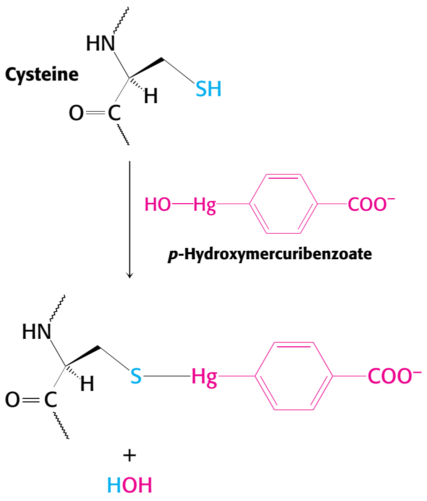


FIGURE 10.4 Modification of cysteine residues. *p*-Hydroxymercuribenzoate reacts with crucial cysteine residues in aspartate transcarbamoylase.

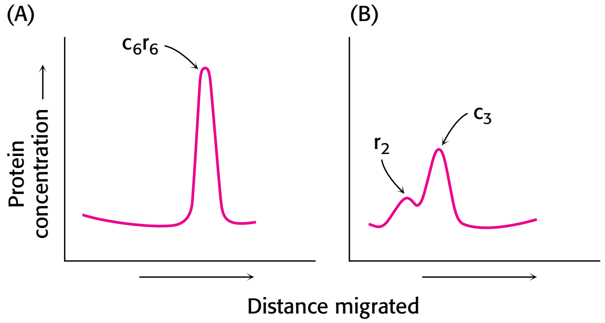


FIGURE 10.5 Ultracentrifugation studies of ATCase. Sedimentation velocity patterns of (A) native ATCase and (B) the enzyme after treatment with *p*-hydroxymercuribenzoate show that the enzyme can be dissociated into regulatory (r) and catalytic (c) subunits. [Data from J. C. Gerhart and H. K. Schachman, *Biochemistry* 4:1054–1062, 1965.]

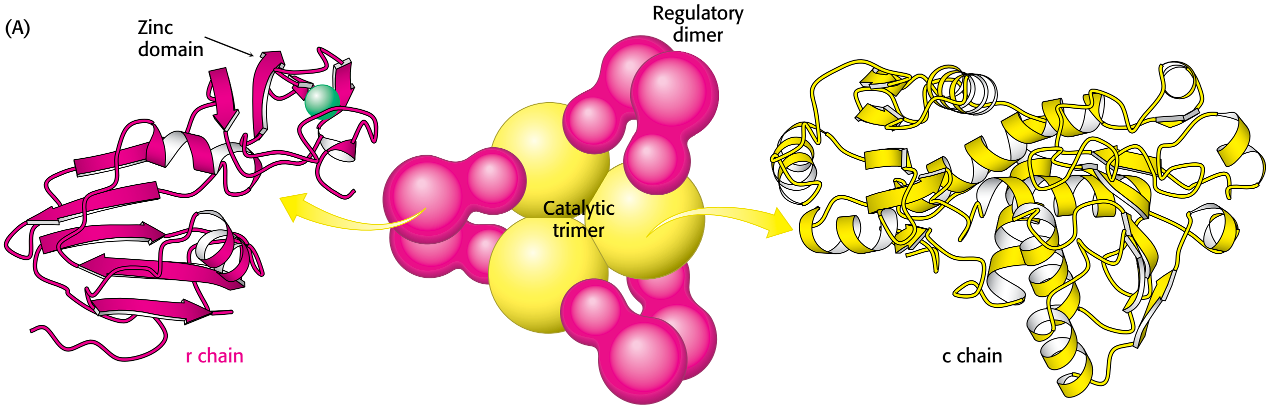
rapidly when they are mixed. The resulting complex has the same structure, c_6r_6, as the native enzyme: two catalytic trimers and three regulatory dimers.

$$2\,c_3 + 3\,r_2 \longrightarrow c_6r_6$$

Most strikingly, the reconstituted enzyme has the same allosteric and kinetic properties as those of the native enzyme. Thus, ATCase is composed of discrete catalytic and regulatory subunits, and *the interaction of the subunits in the native enzyme produces its regulatory and catalytic properties*. The fact that the enzyme can be separated into isolated catalytic and regulatory subunits, which can be reconstituted back to the functional enzyme, allows for a variety of experiments to characterize the allosteric properties of the enzyme (Problems 33 and 34).

Allosteric interactions in ATCase are mediated by large changes in quaternary structure

What are the subunit interactions that account for the properties of ATCase? Significant clues have been provided by the three-dimensional structure of ATCase in various forms. Two catalytic trimers are stacked one on top of the other, linked by three dimers of the regulatory chains (Figure 10.6). There are significant contacts between the catalytic and the regulatory



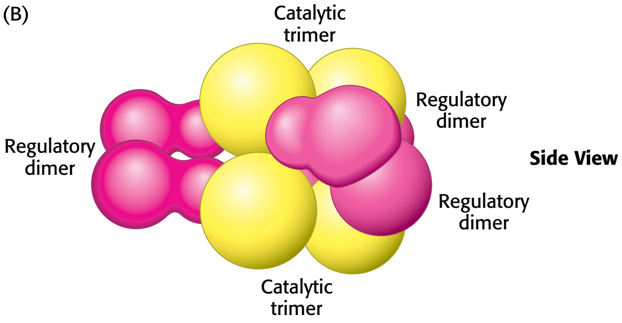


FIGURE 10.6 Structure of ATCase. (A) The quaternary structure of aspartate transcarbamoylase as viewed from the top. The drawing in the center is a simplified representation of the relations between subunits. A single catalytic trimer [catalytic (c) chains, shown in yellow] is visible; in this view, the second trimer is hidden below the one visible. *Notice* that each r chain interacts with a c chain through the zinc domain. (B) A side view of the complex. [Drawn from 1RAI.pdb.]

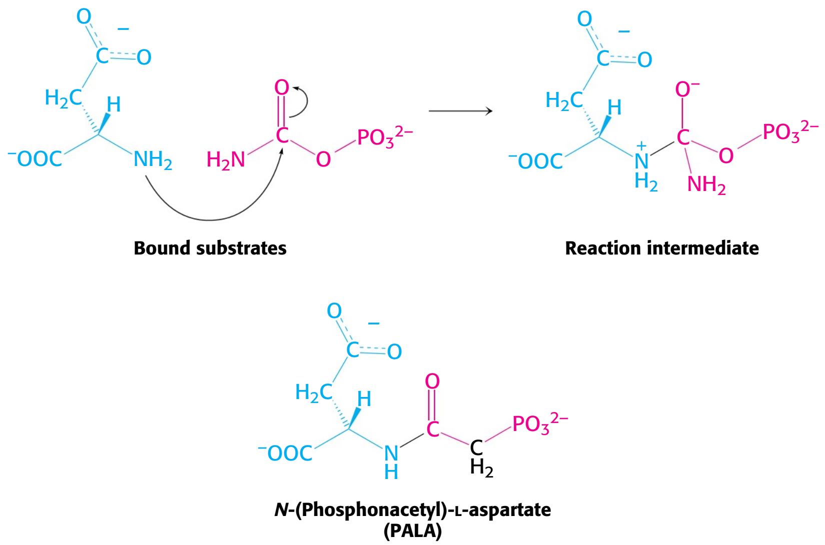


FIGURE 10.7 PALA, a bisubstrate analog. (Top) Nucleophilic attack by the amino group of aspartate on the carbonyl carbon atom of carbamoyl phosphate generates an intermediate on the pathway to the formation of *N*-carbamoylaspartate. (Bottom) *N*-(Phosphonacetyl)-L-aspartate (PALA) is an analog of the reaction intermediate and a potent competitive inhibitor of aspartate transcarbamoylase.

subunits: each r chain within a regulatory dimer interacts with a c chain within a catalytic trimer. The c chain makes contact with a structural domain in the r chain that is stabilized by a zinc ion bound to four cysteine residues. The zinc ion is critical for the interaction of the r chain with the c chain. The mercurial compound *p*-hydroxymercuribenzoate is able to dissociate the catalytic and regulatory subunits because mercury binds strongly to the cysteine residues, displacing the zinc and preventing interaction with the c chain.

To locate the active sites, the enzyme is crystallized in the presence of *N*-(phosphonacetyl)-L-aspartate (PALA), a bisubstrate analog (an analog of the two substrates) that resembles an intermediate along the pathway of catalysis (Figure 10.7). PALA is a potent competitive inhibitor of ATCase that binds to and blocks the active sites. The structure of the ATCase–PALA complex reveals that PALA binds at sites lying at the boundaries between pairs of c chains within a catalytic trimer (Figure 10.8). Each catalytic trimer contributes three active sites to the complete enzyme. Further examination of the ATCase–PALA complex reveals a remarkable change in

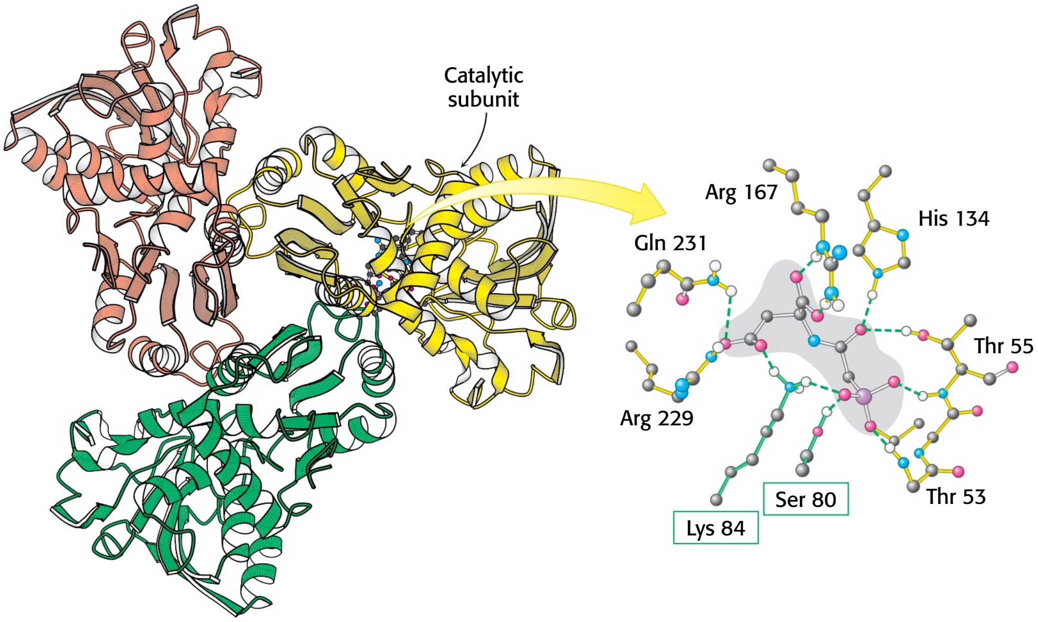


FIGURE 10.8 The active site of ATCase. Some of the crucial active-site residues are shown binding to the inhibitor PALA (shaded gray). *Notice* that the active site is composed mainly of residues from one c chain, but an adjacent c chain also contributes important residues (boxed in green). [Drawn from 8ATC.pdb.]

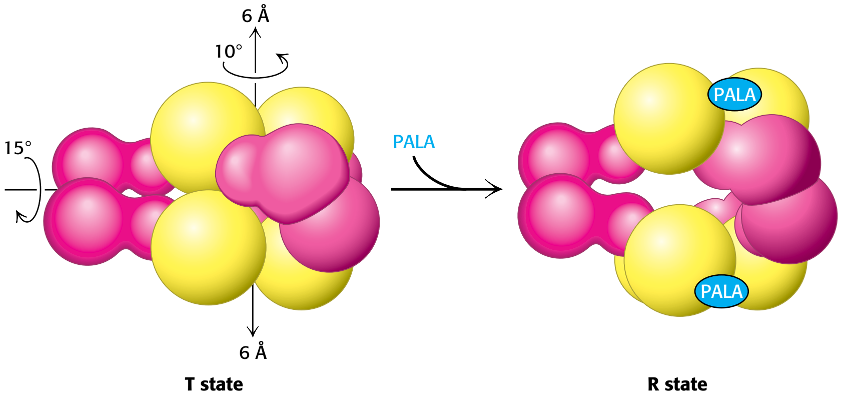


FIGURE 10.9 The T-to-R state transition in ATCase. Aspartate transcarbamoylase exists in two conformations: a compact, relatively inactive form called the tense (T) state and an expanded form called the relaxed (R) state. *Notice* that the structure of ATCase changes dramatically in the transition from the T state to the R State. PALA binding stabilizes the R state.

quaternary structure on binding of PALA. The two catalytic trimers move 12 Å farther apart and rotate approximately 10 degrees about their common threefold axis of symmetry. Moreover, the regulatory dimers rotate approximately 15 degrees to accommodate this motion (Figure 10.9). The enzyme literally expands on PALA binding. In essence, ATCase has two distinct quaternary forms: one that predominates in the absence of substrate or substrate analogs and another that predominates when substrates or analogs are bound. We call these forms the T (for tense) state and the R (for relaxed) state, respectively, as we did for the two quaternary states of hemoglobin.

How can we explain the enzyme's sigmoidal kinetics in light of the structural observations? Like hemoglobin, the enzyme exists in an equilibrium between the T state and the R state.

$$\mathrm{R} \rightleftharpoons \mathrm{T}$$

In the absence of substrate, almost all the enzyme molecules are in the T state because the T state is more stable than the R state. The ratio of the concentration of enzyme in the T state to that in the R state is called the allosteric contstant (L). For most allosteric enzymes, L is on the order of 10^2 to 10^3.

$$\mathrm{L} = \frac{\mathrm{T}}{\mathrm{R}}$$

The T state has a low affinity for substrate and hence shows a low catalytic activity. The occasional binding of a substrate molecule to one active site in an enzyme increases the likelihood that the entire enzyme shifts to the R state with its higher binding affinity. The addition of more substrate has two effects. First, it increases the probability that each enzyme molecule will bind at least one substrate molecule. Second, it increases the average number of substrate molecules bound to each enzyme. The presence of additional substrate will increase the fraction of enzyme molecules in the more active R state because *the position of the equilibrium depends on the number of active sites that are occupied by substrate.* We considered this property, called *cooperativity* because the subunits cooperate with one another, when we discussed the sigmoidal oxygen-binding curve of hemoglobin. The effects of substrates on allosteric enzymes are referred to as *homotropic effects* (from the Greek *homós,* "same").

This mechanism for allosteric regulation is referred to as the *concerted model* because the change in the enzyme is "all or none"; the entire enzyme is converted from T into R, affecting all of the catalytic sites equally. In contrast, the *sequential model* assumes that the binding of ligand to one site on the complex can affect neighboring sites without causing all subunits

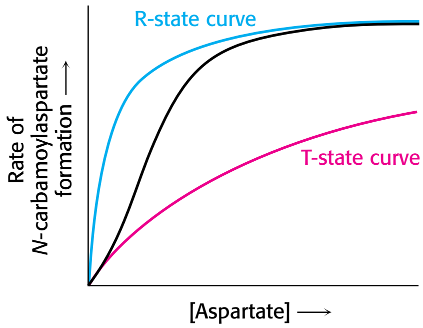


FIGURE 10.10 Basis for the sigmoidal curve. The generation of the sigmoidal curve by the property of cooperativity can be understood by imagining an allosteric enzyme as a mixture of two Michaelis–Menten enzymes, one with a high value of K_M that corresponds to the T state and another with a low value of K_M that corresponds to the R state. As the concentration of substrate is increased, the equilibrium shifts from the T state to the R state, which results in a steep rise in activity with respect to substrate concentration.

to undergo the T-to-R transition (p. 199). Although the concerted model explains the behavior of ATCase well, most other allosteric enzymes have features of both models.

The sigmoidal curve for ATCase can be pictured as a composite of two Michaelis–Menten curves, one corresponding to the less-active T state and the other to the more-active R state. At low concentrations of substrate, the curve closely resembles that of the T state enzyme. As the substrate concentration is increased, the curve progressively shifts to resemble that of the R state enzyme (Figure 10.10).

What is the biochemical advantage of sigmoidal kinetcs? Allosteric enzymes transition from a less active state to a more active state within a narrow range of substrate concentration. The benefit of this behavior is illustrated in Figure 10.11, which compares the kinetics of a Michaelis-Menten enzyme (blue curve) to that of an allosteric enzyme (red curve). In this example, the Michaelis-Menten enzyme requires an approximately 27-fold increase in substrate concentration to increase V_o from 0.1 V_{max} to 0.8 V_{max}. In contrast, the allosteric enzyme requires only about a 4-fold increase in substrate concentration to attain the same increase in velocity. *The activity of allosteric enzymes is more sensitive to changes in substrate concentration near* K_M *than are Michaelis–Menten enzymes with the same* V_{max}. This sensitivity is called a *threshold effect*: below a certain substrate concentration, there is little enzyme activity. However, after the threshold has been reached, enzyme activity increases rapidly. In other words, much like an "on or off" switch, cooperativity ensures that most of the enzyme is either on (R state) or off (T state). The vast majority of allosteric enzymes display sigmoidal kinetics.

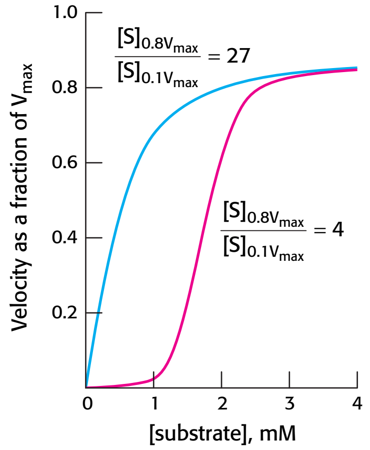


FIGURE 10.11 Allosteric enzymes display threshold effects. As the T-to-R transition occurs, the velocity increases over a narrower range of substrate concentration for an allosteric enzyme (red curve) than for a Michaelis–Menten enzyme (blue curve).

Allosteric regulators modulate the T-to-R equilibrium

We now turn our attention to the effects of pyrimidine nucleotides. As noted earlier, CTP inhibits the action of ATCase. X-ray studies of ATCase in the presence of CTP reveal (1) that the enzyme is in the T state when bound to CTP and (2) that a binding site for this nucleotide exists in each regulatory chain in a domain that does not interact with the catalytic subunit (Figure 10.12). Each active site is more than 50 Å from the nearest CTP-binding site. The question naturally arises, How can CTP inhibit the catalytic activity of the enzyme when it does not interact with the catalytic chain?

The quaternary structural changes observed on substrate-analog binding suggest a mechanism for inhibition by CTP (Figure 10.13). *The binding of the inhibitor CTP to the T state shifts the T-to-R equilibrium in favor of the T state, decreasing net enzyme activity*. CTP increases the allosteric coefficient from 200 in its absence to 1250 when all of the regulatory sites are occupied by CTP. The binding of CTP makes it more difficult for substrate binding to convert the enzyme into the R state.

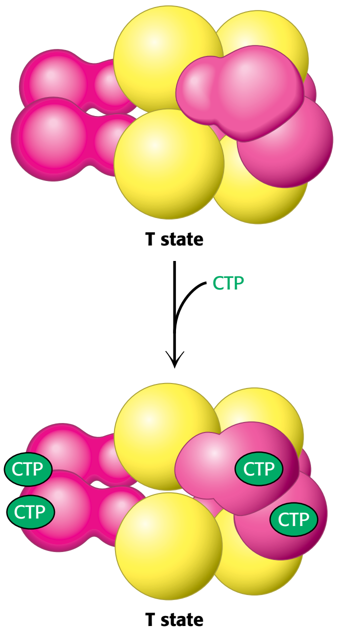


FIGURE 10.12 CTP stabilizes the T state. The binding of CTP to the regulatory subunit of aspartate transcarbamoylase stabilizes the T state.

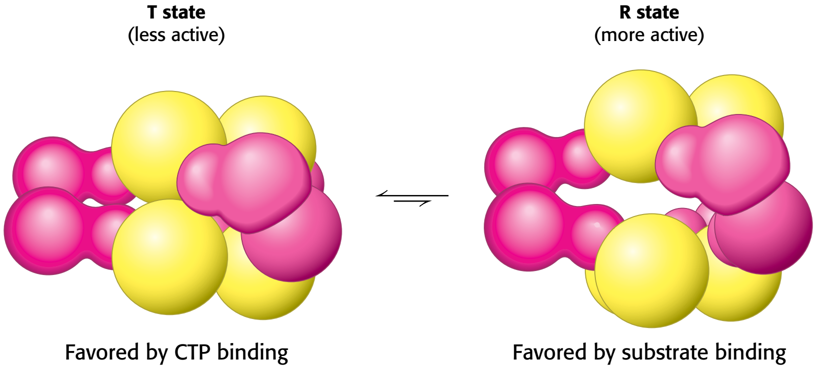


FIGURE 10.13 The R state and the T state are in equilibrium. Even in the absence of any substrate or regulators, aspartate transcarbamoylase exists in equilibrium between the R and the T states. Under these conditions, the T state is favored by a factor of approximately 200.

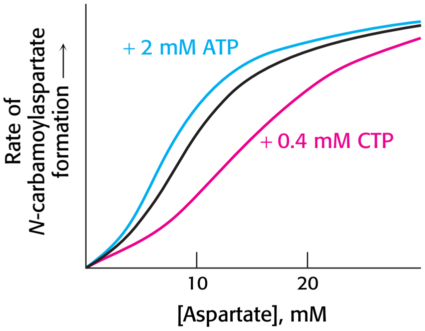


FIGURE 10.14 Effect of CTP and ATP on ATCase kinetics. CTP stabilizes the T state of aspartate transcarbamoylase, making it more difficult for substrate binding to convert the enzyme into the R state. As a result, the curve is shifted to the right, as shown in red. ATP is an allosteric activator of aspartate transcarbamoylase because it stabilizes the R state, making it easier for substrate to bind. As a result, the curve is shifted to the left, as shown in blue.

Consequently, CTP increases the initial phase of the sigmoidal curve (Figure 10.14). More substrate is required to attain a given reaction rate. UTP, the immediate precursor to CTP, also regulates ATCase. While unable to inhibit the enzyme alone, UTP synergistically inhibits ATCase in the presence of CTP.

Interestingly, ATP, too, is an allosteric effector of ATCase, binding to the same site as CTP. However, ATP binding stabilizes the R state, lowering the allosteric coefficient from 200 to 70 and thus *increasing* the reaction rate at a given aspartate concentration (Figure 10.14). At high concentrations of ATP, the kinetic profile shows a less-pronounced sigmoidal behavior. Because ATP and CTP bind at the same site, high levels of ATP prevent CTP from inhibiting the enzyme. The effects of nonsubstrate molecules on allosteric enzymes (such as those of CTP and ATP on ATCase) are referred to as *heterotropic effects* (from the Greek *héteros,* "different"). Substrates generate the sigmoidal curve (homotropic effects), whereas regulators shift the K_M (heterotropic effects). Note, however, that both types of effect are generated by altering the T/R ratio.

The increase in ATCase activity in response to increased ATP concentration has two potential physiological ramifications. First, high ATP concentration signals a high concentration of purine nucleotides in the cell; the increase in ATCase activity will tend to balance the purine and pyrimidine pools. Second, a high concentration of ATP indicates that energy is available for mRNA synthesis and DNA replication and leads to the synthesis of pyrimidines needed for these processes.

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages

Isozymes, or *isoenzymes,* are enzymes that differ in amino acid sequence yet catalyze the same reaction. Typically, these enzymes display different kinetic parameters, such as K_M, or respond to different regulatory molecules. They are encoded by different genes, which usually arise through gene duplication and divergence. Isozymes can often be distinguished from one another by physical properties such as electrophoretic mobility. *Isoform* is a more generic term used when the protein in question is not an enzyme.

The existence of isozymes permits the fine-tuning of metabolism to meet the needs of a given tissue or developmental stage. Consider the example of lactate dehydrogenase (LDH), an enzyme that catalyzes a step in anaerobic

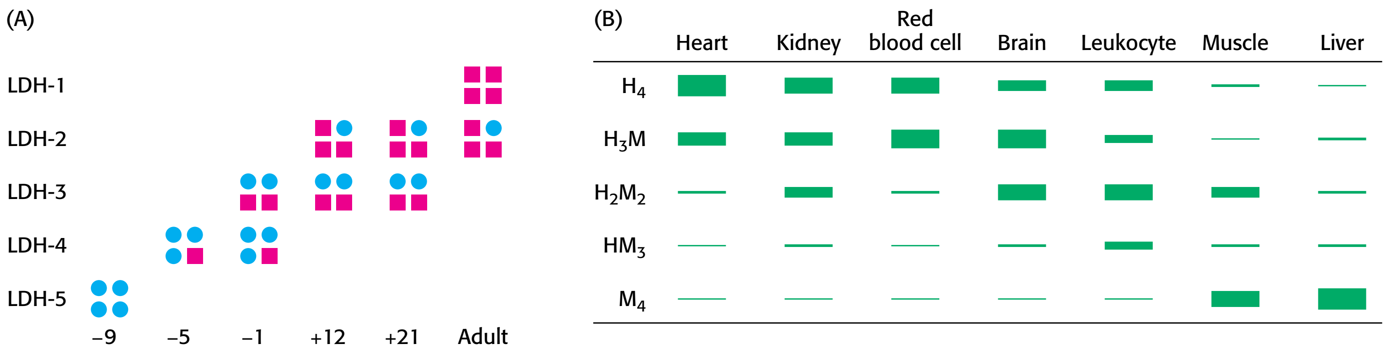


FIGURE 10.15 Isozymes of lactate dehydrogenase. (A) The rat heart lactate dehydrogenase (LDH) isozyme profile changes in the course of development. The H isozyme is represented by squares and the M isozyme by circles. The negative and positive numbers denote the days before and after birth, respectively. (B) LDH isozyme content varies by tissue. [(A) Data from W.-H. Li, *Molecular Evolution* (Sinauer, 1997), p. 283; (B) after K. Urich, *Comparative Animal Biochemistry* (Springer Verlag, 1990), p. 542.]

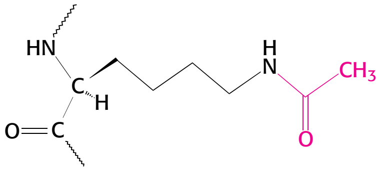
glucose metabolism and glucose synthesis. Human beings have two isozymic polypeptide chains for this enzyme: the H isozyme is highly expressed in heart muscle and the M isozyme is expressed in skeletal muscle. The amino acid sequences are 75% identical. Each functional enzyme is tetrameric, and many different combinations of the two isozymic polypeptide chains are possible. The H_4 isozyme, found in the heart, has a higher affinity for substrates than does the M_4 isozyme. The two isozymes also differ in that high levels of pyruvate allosterically inhibit the H_4 but not the M_4 isozyme. The other combinations, such as H_3M, have intermediate properties. We will consider these isozymes in their biological context in Chapter 16.

The M_4 isozyme functions optimally in the anaerobic environment of hard-working skeletal muscle, whereas the H_4 isozyme does so in the aerobic environment of heart muscle. Indeed, the proportions of these isozymes change throughout the development of the rat heart as the tissue switches from an anaerobic environment to an aerobic one (Figure 10.15A). Figure 10.15B shows the tissue-specific forms of lactate dehydrogenase in adult rat tissues. *Essentially all of the enzymes that we will encounter in later chapters, including allosteric enzymes, exist in isozymic forms.*

The appearance of some isozymes in the blood is a sign of tissue damage, useful for clinical diagnosis. For instance, an increase in serum levels of H_4 relative to H_3M is an indication that a myocardial infarction, or heart attack, has damaged heart-muscle cells, leading to the release of cellular material.

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity

The covalent attachment of a molecule to an enzyme or protein can modify its activity. In these instances, a donor molecule provides the functional moiety being attached. Most modifications are reversible. Phosphorylation and dephosphorylation are common means of covalent modification. The attachment of acetyl groups to lysine residues by acetyltransferases and their removal by deacetylases are another example. Histones—proteins that are packaged with DNA into chromosomes—are extensively acetylated and deacetylated in vivo on lysine residues (Section 31.3). More heavily acetylated histones are associated with genes that are being actively transcribed. Although protein acetylation was originally discovered as a modification of histones, we now know that it is a major means of regulation, with more than 2000 different proteins in mammalian cells regulated by acetylation.



Acetylated lysine

TABLE 10.1 Common covalent modifications of protein activity

Modification	Donor molecule	Example of modified protein	Protein function
Phosphorylation	ATP	Glycogen phosphorylase	Glucose homeostasis; energy transduction
Acetylation	Acetyl CoA	Histones	DNA packing; transcription
Myristoylation	Myristoyl CoA	Src	Signal transduction
ADP ribosylation	NAD^+	RNA polymerase	Transcription
Farnesylation	Farnesyl pyrophosphate	Ras	Signal transduction
γ-Carboxylation	HCO_3^-	Thrombin	Blood clotting
Sulfation	3′-Phosphoadenosine-5′-phosphosulfate	Fibrinogen	Blood-clot formation
Ubiquitination	Ubiquitin	Cyclin	Control of cell cycle

Protein acetylation appears to be especially important in the regulation of metabolism. The acetyltransferase and deacetylase enzymes are themselves regulated by phosphorylation, showing that the covalent modification of a protein can be controlled by the covalent modification of the modifying enzymes.

Modification is not readily reversible in some cases. The irreversible attachment of a lipid group causes some proteins in signal-transduction pathways, such as Ras (a GTPase) and Src (a protein tyrosine kinase), to become affixed to the cytoplasmic face of the plasma membrane. Fixed in this location, the proteins are better able to receive and transmit information that is being passed along their signaling pathways (Chapter 14). Mutations in both Ras and Src are seen in a wide array of cancers. The attachment of the small protein ubiquitin can signal that a protein is to be destroyed, the ultimate means of regulation (Chapter 23). The protein cyclin must be ubiquitinated and destroyed before a cell can enter anaphase and proceed through the cell cycle.

Virtually all the metabolic processes that we will examine are regulated in part by covalent modification. Indeed, the allosteric properties of many enzymes are altered by covalent modification. Table 10.1 lists some of the common covalent modifications.

Kinases and phosphatases control the extent of protein phosphorylation

We will see phosphorylation used as a regulatory mechanism in virtually every metabolic process in eukaryotic cells. Indeed, as much as 30% of eukaryotic proteins are phosphorylated. The enzymes catalyzing phosphorylation reactions are called *protein kinases*. These enzymes constitute one of the largest protein families known: there are more than 500 homologous kinases in human beings. This multiplicity of enzymes allows regulation to be fine-tuned according to a specific tissue, time, or substrate.

ATP is the most common donor of phosphoryl groups. The terminal (γ) phosphoryl group of ATP is transferred to a specific amino acid of the acceptor protein or enzyme. In eukaryotes, the acceptor residue is commonly one of the three containing a hydroxyl group in its side chain. Transfers to *serine* and *threonine* residues are handled by one class of protein kinases and to *tyrosine* residues by another. Tyrosine kinases, which are unique to multicellular organisms, play pivotal roles in growth regulation, and mutations in these enzymes are commonly observed in cancer cells.

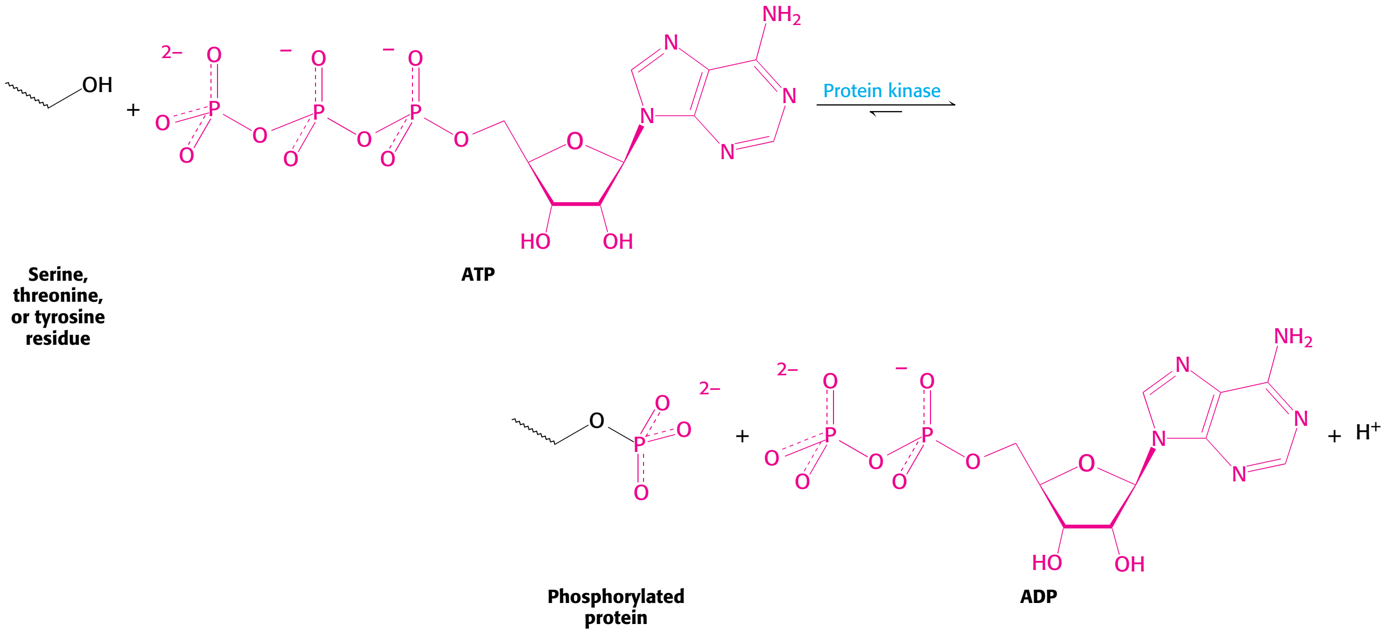


Table 10.2 lists a few of the known serine and threonine protein kinases. The acceptors in protein-phosphorylation reactions are located inside cells, where the phosphoryl-group donor ATP is abundant. Proteins that are entirely extracellular are not regulated by reversible phosphorylation.

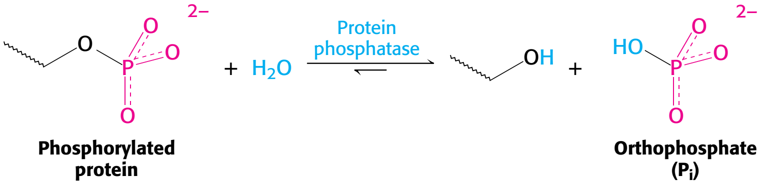
Protein kinases vary in their degree of specificity. *Dedicated protein kinases* phosphorylate a single protein or several closely related ones. *Multifunctional protein kinases* modify many different targets; they have a wide reach and can coordinate diverse processes. Comparisons of amino acid sequences of many phosphorylation sites show that a multifunctional kinase recognizes related sequences. For example, the *consensus sequence* recognized by protein kinase A is Arg-Arg-X-*Ser*-Z or Arg-Arg-X-*Thr*-Z, in which X is a small residue, Z is a large hydrophobic one, and *Ser* or *Thr* is the site of phosphorylation. However, this sequence is not absolutely required. Lysine, for example, can substitute for one of the arginine residues but with some loss of affinity. Thus, *the primary determinant of specificity is the amino acid sequence surrounding the serine or threonine phosphorylation site.* However, distant residues can contribute to specificity. For instance, a change in protein conformation can open or close access to a possible phosphorylation site.

Protein phosphatases reverse the effects of kinases by catalyzing the removal of phosphoryl groups attached to proteins. The enzyme hydrolyzes the bond attaching the phosphoryl group.

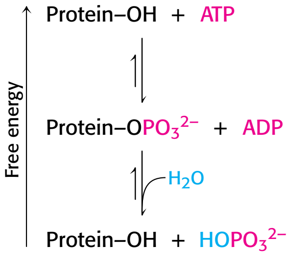
TABLE 10.2 Examples of serine and threonine kinases and their activating signals

Signal	Enzyme
Cyclic nucleotides	Cyclic AMP-dependent protein kinase Cyclic GMP-dependent protein kinase
Ca^{2+} and calmodulin	Ca^{2+}–calmodulin protein kinase Phosphorylase kinase or glycogen synthase kinase 2
AMP	AMP-activated kinase
Diacylglycerol	Protein kinase C
Metabolic intermediates and other "local" effectors	Many target-specific enzymes, such as pyruvate dehydrogenase kinase and branched-chain ketoacid dehydrogenase kinase

Source: Information from D. Fell, *Understanding the Control of Metabolism* (Portland Press, 1997), Table 7.2.



The unmodified hydroxyl-containing side chain is regenerated and orthophosphate (P_i) is produced. This family of enzymes, of which there are about 200 members in human beings, plays a vital role in cells because the enzymes turn off the signaling pathways that are activated by kinases. One class of highly conserved phosphatase called PP2A suppresses the cancer-promoting activity of certain kinases.



Importantly, the phosphorylation and dephosphorylation reactions are not the reverse of one another; each is essentially irreversible under physiological conditions. Furthermore, both reactions take place at negligible rates in the absence of enzymes. Thus, phosphorylation of a protein substrate will take place only through the action of a specific protein kinase and at the expense of ATP cleavage, and dephosphorylation will take place only through the action of a phosphatase. The result is that target proteins cycle unidirectionally between unphosphorylated and phosphorylated forms. The rate of cycling between the phosphorylated and the dephosphorylated states depends on the relative activities of kinases and phosphatases.

Phosphorylation is a highly effective means of regulating the activities of target proteins

Phosphorylation is a common covalent modification of proteins in all forms of life, which leads to the question, What makes protein phosphorylation so valuable in regulating protein function that its use is ubiquitous? Phosphorylation is a highly effective means of controlling the activity of proteins for several reasons:

1. *The free energy of phosphorylation is large.* Of the -50 kJ mol^{-1} (-12 kcal mol^{-1}) provided by ATP, about half is consumed in making phosphorylation irreversible; the other half is conserved in the phosphorylated protein. A free-energy change of 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}) corresponds to a factor of 10 in an equilibrium constant. Hence, phosphorylation can change the conformational equilibrium between different functional states by a large factor, of the order of 10^4. In essence, the energy expenditure allows for a stark shift from one conformation to another.

2. *A phosphoryl group adds two negative charges to a modified protein.* These new charges may disrupt electrostatic interactions in the unmodified protein and allow new electrostatic interactions to be formed. Such structural changes can markedly alter substrate binding and catalytic activity.

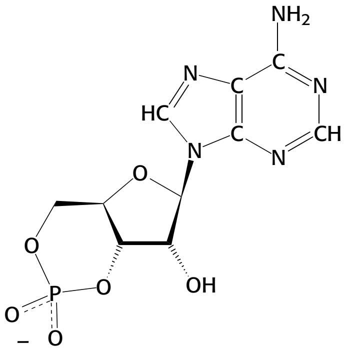
3. *A phosphoryl group can form three or more hydrogen bonds.* The tetrahedral geometry of a phosphoryl group makes these bonds highly directional, allowing for specific interactions with hydrogen-bond donors.

4. *Phosphorylation and dephosphorylation can take place in less than a second or over a span of hours.* The kinetics can be adjusted to meet the timing needs of a physiological process.

5. *Phosphorylation often evokes highly amplified effects.* A single activated kinase can phosphorylate hundreds of target proteins in a short interval. If the target protein is an enzyme, it can in turn transform a large number of substrate molecules.

6. *ATP is the cellular energy currency* (Chapter 15). The use of this compound as a phosphoryl-group donor links the energy status of the cell to the regulation of metabolism.

Cyclic AMP activates protein kinase A by altering the quaternary structure



Cyclic adenosine monophosphate (cAMP)

Let us examine a specific protein kinase that helps animals cope with stressful situations. The "flight or fight" response is common to many animals presented with a dangerous or exciting situation. Muscle becomes primed for action. This priming is the result of the activity of a particular protein kinase. In this case, the hormone epinephrine (adrenaline) triggers the formation of cyclic AMP (cAMP), an intracellular messenger formed by the cyclization of ATP. Cyclic AMP subsequently activates a key enzyme: *protein kinase A* (PKA). The kinase alters the activities of target proteins by phosphorylating specific serine or threonine residues. The striking finding is that *most effects of cAMP in eukaryotic cells are achieved through the activation of PKA by cAMP.*

PKA provides a clear example of the integration of allosteric regulation and phosphorylation. PKA is activated by cAMP concentrations near 10 nM. The quaternary structure is reminiscent of that of ATCase. Like that enzyme, PKA in muscle consists of two kinds of subunits: a 49-kDa regulatory (R) subunit and a 38-kDa catalytic (C) subunit. In the absence of cAMP, the regulatory and catalytic subunits form an R_2C_2 complex that is enzymatically inactive (Figure 10.16). The binding of two molecules of cAMP to each of the regulatory subunits leads to the dissociation of R_2C_2 into an R_2 subunit and two C subunits. These free catalytic subunits are enzymatically active. Thus, *the binding of cAMP to the regulatory subunit relieves its inhibition of the catalytic subunit.* PKA, like most other kinases, exists in isozymic forms for fine-tuning regulation to meet the needs of a specific cell or developmental stage. In mammals, four isoforms of the R subunit and three of the C subunit are encoded in the genome.

How does the binding of cAMP activate the kinase? Each R chain contains the sequence Arg-Arg-Gly-*Ala*-Ile, which matches the consensus sequence for phosphorylation except for the presence of alanine in place of serine. In the R_2C_2 complex, this *pseudosubstrate sequence* of R occupies the catalytic site of C, thereby preventing the entry of protein substrates (Figure 10.16). The binding of cAMP to the R chains allosterically moves the pseudosubstrate sequences out of the catalytic sites. The released C chains are then free to bind and phosphorylate substrate proteins. Interestingly, the cAMP-binding domain of the R subunit is highly conserved and found in all organisms.

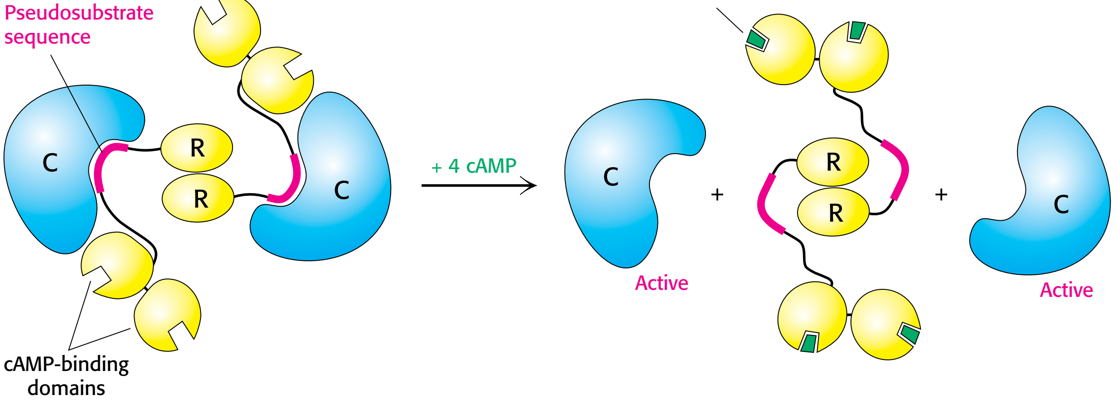


FIGURE 10.16 Regulation of protein kinase A. The binding of four molecules of cAMP activates protein kinase A by dissociating the inhibited holoenzyme (R_2C_2) into a regulatory subunit (R_2) and two catalytically active subunits (C). Each R chain includes cAMP-binding domains and a pseudosubstrate sequence.

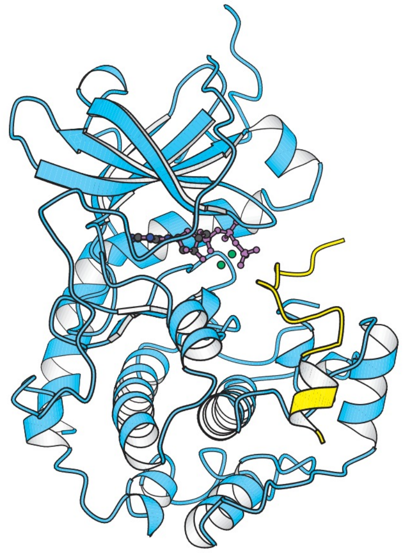


FIGURE 10.17 **Protein kinase A bound to an inhibitor.** This ribbon model shows a complex of the catalytic subunit of protein kinase A with an inhibitor (yellow) bearing a pseudosubstrate sequence. *Notice* that the inhibitor binds to the active site, a cleft between the domains of the enzyme. The bound ATP (purple)-Mn^{2+} (green) is in the active site adjacent to the site to which the inhibitor is bound. [Drawn from 1ATP.pdb.]

ATP and the target protein bind to a deep cleft in the catalytic subunit of protein kinase A

X-ray crystallography reveals the three-dimensional structure of the catalytic subunit of PKA bound to ATP and a 20-residue peptide inhibitor. The 350-residue catalytic subunit of PKA has two lobes, an amino terminal N-lobe and a carboxyl terminal C-lobe (Figure 10.17). ATP and part of the inhibitor fill a deep cleft between the lobes. The N-lobe makes many contacts with ATP–Mn^{2+}, whereas the C-lobe binds the peptide and contributes the key catalytic residues. As with other kinases, the two lobes move closer to one another on substrate binding; mechanisms that restrict this domain closure provide a means of regulating protein kinase activity. The PKA structure has broad significance because residues 40 to 280 constitute a conserved catalytic core, called the *kinase fold*, that is common to essentially all known protein kinases. We see here an example of a successful biochemical solution to a problem (in this case, protein phosphorylation) being employed many times in the course of evolution.

The bound peptide in this crystal occupies the active site because it contains the pseudosubstrate sequence Arg-Arg-Asn-*Ala*-Ile (Figure 10.18). The structure of the complex reveals the interactions by which the enzyme recognizes the consensus sequence. The guanidinium group of the first arginine residue forms an ion pair with the carboxylate side chain of a glutamate residue (Glu 127) of the enzyme. The second arginine likewise interacts with two other carboxylate groups. The nonpolar side chain of isoleucine, which matches Z in the consensus sequence (p. 295), fits snugly in a hydrophobic groove formed by two leucine residues of the enzyme.

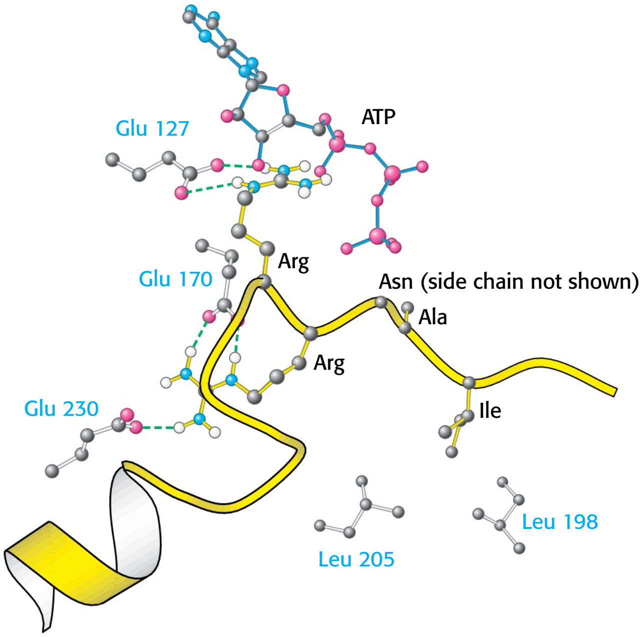


FIGURE 10.18 **Binding of pseudosubstrate to protein kinase A.** *Notice* that the inhibitor makes multiple contacts with the enzyme. The two arginine side chains of the pseudosubstrate ionically interact with three glutamate carboxylate groups (green dashed lines). Hydrophobic interactions also are important in the recognition of substrate. The isoleucine residue of the pseudosubstrate is in contact with a pair of leucine residues of the enzyme.

10.4 Many Enzymes are Activated by Specific Proteolytic Cleavage

We turn now to a different mechanism of enzyme regulation. Many enzymes acquire full enzymatic activity as they spontaneously fold into their characteristic three-dimensional forms. In contrast, the folded forms of other enzymes are inactive until the cleavage of one or a few specific peptide bonds. The inactive precursor is called a *zymogen* or a *proenzyme*. An energy source such as ATP is not needed for cleavage. Therefore, in contrast with reversible regulation by phosphorylation, even proteins located outside cells can be activated by this means. Another noteworthy difference is that proteolytic activation, in contrast with allosteric control and reversible covalent modification, is irreversible, taking place just once in the life of an enzyme molecule.

Specific proteolysis is a common means of activating enzymes and other proteins in biological systems. For example:

1. The *digestive enzymes* that hydrolyze foodstuffs are synthesized as zymogens in the stomach and pancreas (Table 10.3).

TABLE 10.3 Gastric and pancreatic zymogens

Site of synthesis	Zymogen	Active enzyme
Stomach	Pepsinogen	Pepsin
Pancreas	Chymotrypsinogen	Chymotrypsin
Pancreas	Trypsinogen	Trypsin
Pancreas	Procarboxypeptidase	Carboxypeptidase

2. *Blood clotting* is mediated by a cascade of proteolytic activations that ensures a rapid and amplified response to trauma.

3. Some protein hormones are synthesized as inactive precursors. For example, *insulin* is derived from *proinsulin* by proteolytic removal of a peptide.

4. The fibrous protein *collagen,* the major constituent of skin and bone, is derived from *procollagen,* a soluble precursor.

5. Many *developmental processes* are controlled by the activation of zymogens. For example, in the metamorphosis of a tadpole into a frog, large amounts of collagen are resorbed from the tail in the course of a few days. Likewise, much collagen is broken down in a mammalian uterus after delivery. The conversion of *procollagenase* into *collagenase,* the active protease responsible for collagen breakdown, is precisely timed in these remodeling processes.

6. *Programmed cell death,* or *apoptosis,* is mediated by proteolytic enzymes called *caspases,* which are synthesized in precursor form as *procaspases.* When activated by various signals, caspases function to cause cell death in most organisms, ranging from *C. elegans* to human beings. Apoptosis provides a means of sculpting the shapes of body parts in the course of development and a means of eliminating damaged or infected cells.

We next examine the activation and control of zymogens, using enzymes responsible for digestion and blood-clot formation as examples.

Chymotrypsinogen is activated by specific cleavage of a single peptide bond

Chymotrypsin is a digestive enzyme that hydrolyzes proteins. Chymotrypsin, whose mechanism of action was described in detail in Chapter 9, specifically cleaves peptide bonds on the carboxyl side of amino acid residues with large, hydrophobic R groups (Table 8.6). Its inactive precursor, *chymotrypsinogen,*

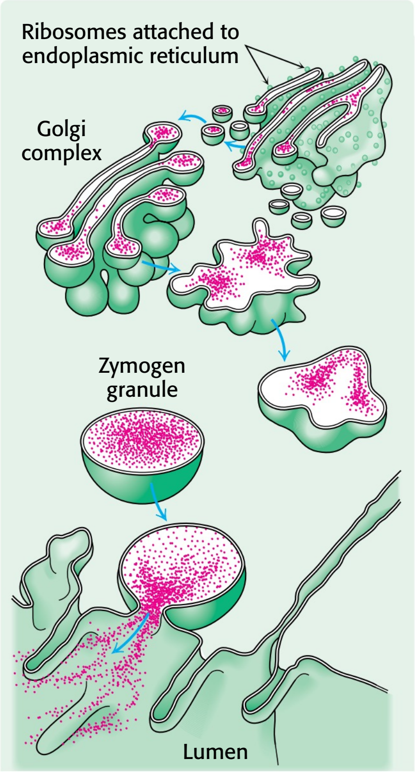

Ribosomes attached to endoplasmic reticulum
Golgi complex
Zymogen granule
Lumen

FIGURE 10.19 Secretion of zymogens by an acinar cell of the pancreas. Zymogens are synthesized on ribosomes attached to the endoplasmic reticulum. They are subsequently processed in the Golgi apparatus and packaged into zymogen or secretory granules. With the proper signal, the granules fuse with the plasma membrane, discharging their contents into the lumen of the pancreatic ducts. Cell cytoplasm is depicted as pale green. Membranes and lumen are shown as dark green.

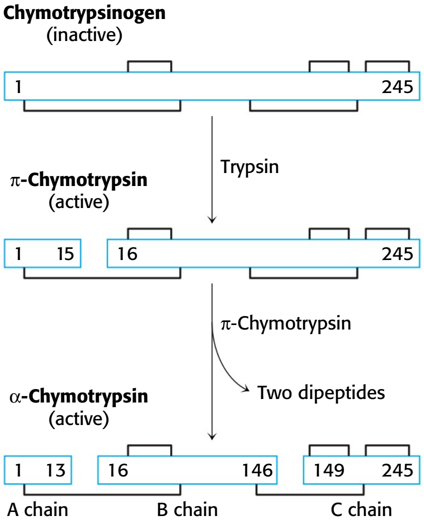


FIGURE 10.20 Proteolytic activation of chymotrypsinogen. The three chains of α-chymotrypsin are linked by two interchain disulfide bonds (A to B, and B to C). The approximate positions of disulfide bonds are shown.

is synthesized in the pancreas, as are several other zymogens and digestive enzymes. Indeed, the pancreas is one of the most active organs in synthesizing and secreting proteins. The enzymes and zymogens are synthesized in the acinar cells of the pancreas and stored inside membrane-bounded granules (Figure 10.19). The zymogen granules accumulate at the apex of the acinar cell; when the cell is stimulated by a hormonal signal or a nerve impulse, the contents of the granules are released into a duct leading into the duodenum.

Chymotrypsinogen, a single polypeptide chain consisting of 245 amino acid residues, is virtually devoid of enzymatic activity. It is converted into a fully active enzyme when the peptide bond joining arginine 15 and isoleucine 16 is cleaved by trypsin (Figure 10.20). The resulting active enzyme, called π-chymotrypsin, then acts on other π-chymotrypsin molecules by removing two dipeptides to yield α-chymotrypsin, the stable form of the enzyme. The three resulting chains in α-chymotrypsin remain linked to one another by two interchain disulfide bonds. The striking feature of this activation process is that *cleavage of a single specific peptide bond transforms the protein from a catalytically inactive form into one that is fully active.*

Proteolytic activation of chymotrypsinogen leads to the formation of a substrate-binding site

How does cleavage of a single peptide bond activate the zymogen? The cleavage of the peptide bond between amino acids 15 and 16 triggers key conformational changes, which were revealed by the elucidation of the three-dimensional structure of chymotrypsinogen.

1. The newly formed *amino-terminal group of isoleucine 16 turns inward and forms an ionic bond with aspartate 194* in the interior of the chymotrypsin molecule (Figure 10.21).

2. This electrostatic interaction triggers a number of conformational changes. Methionine 192 moves from a deeply buried position in the zymogen to the surface of the active enzyme, and residues 187 and 193 move farther apart from

each other. These changes result in the formation of the *substrate-specificity site* for aromatic and bulky nonpolar groups. One side of this site is made up of residues 189 through 192. *This cavity for binding part of the substrate is not fully formed in the zymogen.*

3. The tetrahedral transition state generated by chymotrypsin has an oxyanion (a negatively charged carbonyl oxygen atom) that is stabilized by hydrogen bonds with two NH groups of the main chain of the enzyme (Figure 9.9). One of these NH groups is not appropriately located in chymotrypsinogen, and so the site stabilizing the oxyanion (the oxyanion hole, p. 257) is incomplete in the zymogen.

4. The conformational changes elsewhere in the molecule are very small. Thus, *the switching on of enzymatic activity in a protein can be accomplished by discrete, highly localized conformational changes that are triggered by the hydrolysis of a single peptide bond.*

The generation of trypsin from trypsinogen leads to the activation of other zymogens

The structural changes accompanying the activation of *trypsinogen,* the precursor of the proteolytic enzyme *trypsin,* are different from those in the activation of chymotrypsinogen. Four regions of the polypeptide are very flexible in the zymogen, whereas they have a well-defined conformation in trypsin. The resulting structural changes also complete the formation of the oxyanion hole.

The digestion of proteins and other molecules in the duodenum requires the concurrent action of several enzymes, because each is specific for a limited number of side chains. Thus, the zymogens must be switched on at the same time. Coordinated control is achieved by the action of *trypsin as the common activator of all the pancreatic zymogens*—trypsinogen, chymotrypsinogen, proelastase, procarboxypeptidase, and prolipase, the inactive precursor of a lipid-degrading enzyme. To produce active trypsin, the cells that line the duodenum display a membrane-embedded enzyme, *enteropeptidase,* which hydrolyzes a unique lysine–isoleucine peptide bond in trypsinogen as the zymogen enters the duodenum from the pancreas. The small amount of trypsin produced in this way activates more trypsinogen and the other zymogens (Figure 10.22). Thus, *the formation of trypsin by enteropeptidase is the master activation step.*

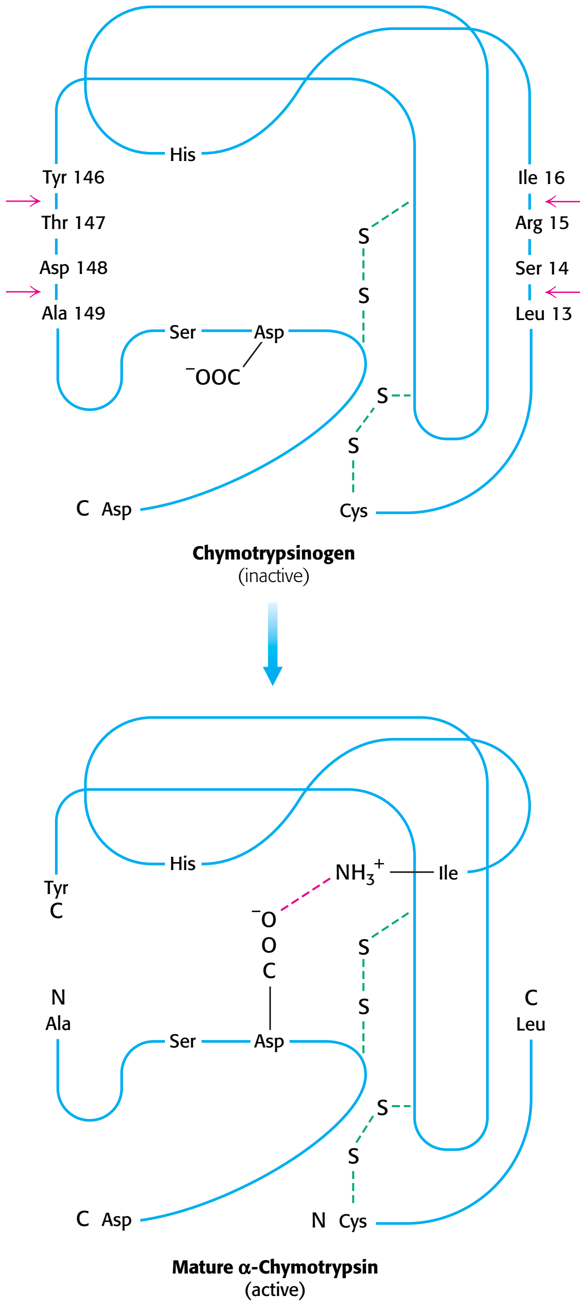

His
Tyr 146
Thr 147
Asp 148
Ala 149
Ile 16
Arg 15
Ser 14
Leu 13
S
S
S
S
Ser
Asp
⁻OOC
C Asp
Cys
Chymotrypsinogen
(inactive)
His
Tyr
C
NH3+
Ile
⁻O
O
C
S
S
S
S
N
Ala
C
Leu
Ser
Asp
C Asp
N Cys
Mature α-Chymotrypsin
(active)

FIGURE 10.21 Conformations of chymotrypsinogen and chymotrypsin. The electrostatic interaction between the α-amino group of isoleucine 16 and the carboxylate of aspartate 194, essential for the structure of active chymotrypsin, is possible only only after cleavage of the peptide bond between isoleucine and arginine in chymotrypsinogen. [Information from Gregory A. Petsko and Dagmar Ringe, *Protein Structure and Function* (Sinauer, 2003), p. 3-16, Figure 3-31.]

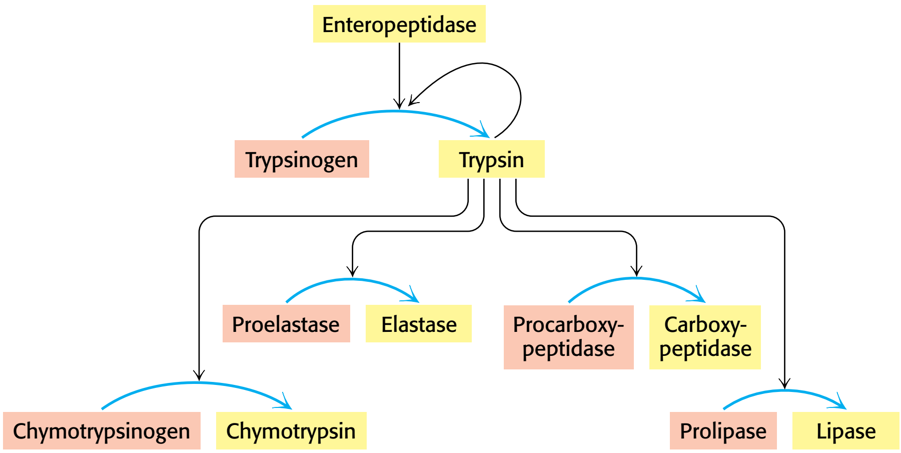

Enteropeptidase
Trypsinogen
Trypsin
Proelastase
Elastase
Procarboxy-peptidase
Carboxy-peptidase
Chymotrypsinogen
Chymotrypsin
Prolipase
Lipase

FIGURE 10.22 Zymogen activation by proteolytic cleavage. Enteropeptidase initiates the activation of the pancreatic zymogens by activating trypsin, which then activates other zymogens. Active enzymes are shown in yellow; zymogens are shown in orange.

Some proteolytic enzymes have specific inhibitors

The conversion of a zymogen into a protease by cleavage of a single peptide bond is a precise means of switching on enzymatic activity. However, this activation step is irreversible, and so a different mechanism is needed to terminate proteolysis. Specific protease inhibitors accomplish this task. *Serpins*, *ser*ine protease *in*hibitors, are an example of one such family of inhibitors. For example, *pancreatic trypsin inhibitor,* a 6-kDa protein, inhibits trypsin by binding very tightly to its active site. The dissociation constant of the complex is 0.1 pM, which corresponds to a standard free energy of binding of about −75 kJ mol^{-1} (−18 kcal mol^{-1}). In contrast with nearly all known protein assemblies, this complex is not dissociated into its constituent chains by treatment with denaturing agents such as 8 M urea or 6 M guanidine hydrochloride (p. 47).

The reason for the exceptional stability of the complex is that pancreatic trypsin inhibitor is a very effective substrate analog. X-ray analyses show that the inhibitor lies in the active site of the enzyme, positioned such that the side chain of lysine 15 of this inhibitor interacts with the aspartate side chain in the specificity pocket of trypsin. In addition, there are many hydrogen bonds between the main chain of trypsin and that of its inhibitor. Furthermore, the carbonyl group of lysine 15 and the surrounding atoms of the inhibitor fit snugly in the active site of the enzyme. Comparison of the structure of the inhibitor bound to the enzyme with that of the free inhibitor reveals that *the structure is essentially unchanged on binding to the enzyme* (Figure 10.23). Thus, the inhibitor is preorganized into a structure that is highly complementary to the enzyme's active site. Indeed, the peptide bond between lysine 15 and alanine 16 in pancreatic trypsin inhibitor is cleaved but at a very slow rate: the half-life of the trypsin–inhibitor complex is several months. In essence, the inhibitor is a substrate, but its intrinsic structure is so nicely complementary to the enzyme's active site that it binds very tightly, rarely progressing to the transition state and is turned over slowly.

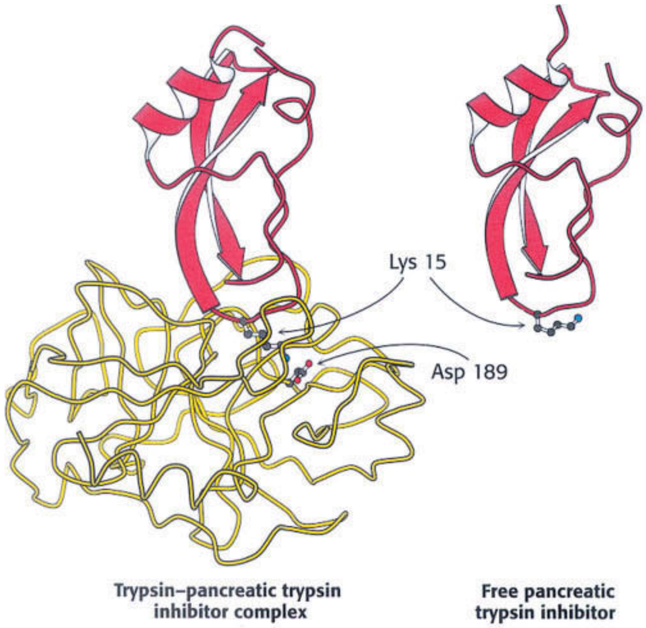

Lys 15
Asp 189
Trypsin–pancreatic trypsin inhibitor complex
Free pancreatic trypsin inhibitor

FIGURE 10.23 Interaction of trypsin with its inhibitor. Structure of a complex of trypsin (yellow) and pancreatic trypsin inhibitor (red). *Notice* that lysine 15 of the inhibitor penetrates into the active site of the enzyme. There it forms an ionic bond with aspartate 189 in the active site. *Also notice* that bound inhibitor and the free inhibitor are almost identical in structure. [Drawn from 1BPI.pdb.]

The amount of trypsin is much greater than the amount of inhibitor. Why does trypsin inhibitor exist? Recall that trypsin activates other zymogens. Consequently, the prevention of even small amounts of trypsin from initiating the cascade while the zymogens are still in the pancreas or pancreatic ducts is vital. Trypsin inhibitor binds to any prematurely activated trypsin molecules in the pancreas or pancreatic ducts. This inhibition prevents severe damage to those tissues, which could lead to acute pancreatitis.

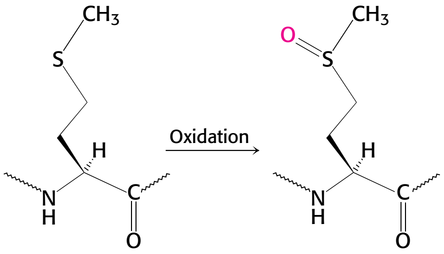


FIGURE 10.24 Oxidation of methionine to methionine sulfoxide.

Pancreatic trypsin inhibitor is not the only important protease inhibitor. A 53-kDa plasma protein, *α_1-antitrypsin* (also called *α_1-antiproteinase*), protects tissues from digestion by elastase, a secretory product of neutrophils (white blood cells that engulf bacteria). *Antielastase* would be a more accurate name for this inhibitor, because it blocks *elastase* much more effectively than it blocks trypsin. Like pancreatic trypsin inhibitor, α_1-antitrypsin blocks the action of target enzymes by binding nearly irreversibly to their active sites. Genetic disorders leading to a deficiency of α_1-antitrypsin illustrate the physiological importance of this inhibitor. For example, the substitution of lysine for glutamate at residue 53 in the type Z mutant slows the secretion of this inhibitor from liver cells. Serum levels of the inhibitor are about 15% of normal in people homozygous for this defect. The consequence is that excess elastase destroys alveolar walls in the lungs by digesting elastic fibers and other connective-tissue proteins.

The resulting clinical condition is called *emphysema* (also known as *chronic obstructive pulmonary disease* [COPD]). People with emphysema must breathe much harder than normal people to exchange the same volume of air because their alveoli are much less resilient than normal. Cigarette smoking markedly increases the likelihood that even a type Z heterozygote will develop emphysema. The reason is that smoke oxidizes methionine 358 of the inhibitor (Figure 10.24), a residue essential for binding elastase. Indeed, this methionine side chain is the bait that selectively traps elastase. The *methionine sulfoxide* oxidation product, in contrast, does not lure elastase, a striking consequence of the insertion of just one oxygen atom into a protein and a remarkable example of the effect of human behavior on biochemistry. We will consider another protease inhibitor, antithrombin III, when we examine the control of blood clotting.

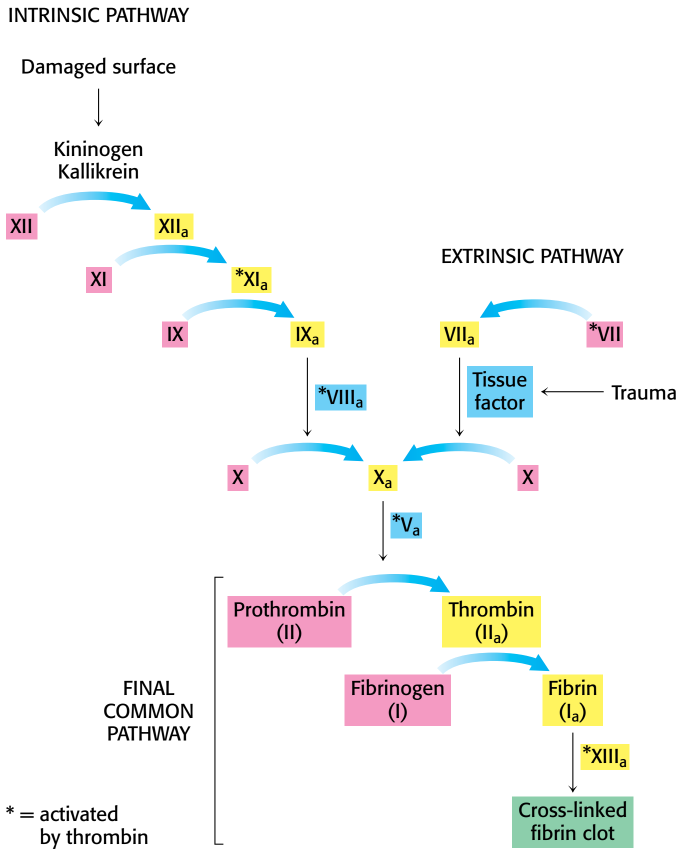
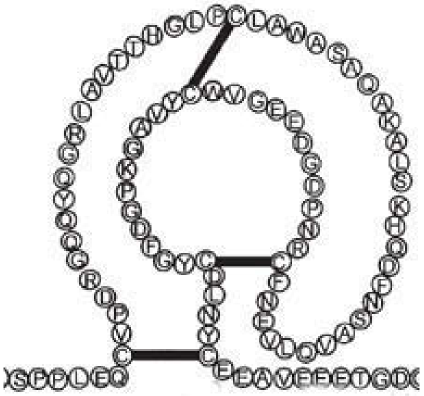


FIGURE 10.25 Blood-clotting cascade. A fibrin clot is formed by the interplay of the intrinsic, extrinsic, and final common pathways. The intrinsic pathway begins with the activation of factor XII (Hageman factor) by contact with abnormal surfaces produced by injury. The extrinsic pathway is triggered by trauma, which releases tissue factor (TF). TF forms a complex with VII, which initiates a cascade-activating thrombin. Inactive forms of clotting factors are shown in red; their activated counterparts (indicated by the subscript "a") are in yellow. Stimulatory proteins that are not themselves enzymes are shown in blue boxes. A striking feature of this process is that the activated form of one clotting factor catalyzes the activation of the next factor.

Blood clotting is accomplished by a cascade of zymogen activations

Enzymatic cascades are often employed in biochemical systems to achieve a rapid response. In a cascade, an initial signal institutes a series of steps, each of which is catalyzed by an enzyme. At each step, the signal is amplified. For instance, if a signal molecule activates an enzyme that in turn activates 10 enzymes and each of the 10 enzymes in turn activates 10 additional enzymes, after four steps the original signal will have been amplified 10,000-fold. *Hemostasis,* the process of blood clot formation and dissolution, requires a *cascade of zymogen activations:* the activated form of one clotting factor catalyzes the activation of the next (Figure 10.25). Thus, very small amounts of the initial factors suffice to trigger the cascade, ensuring a rapid response to trauma.



The second kringle domain of protrhombin

Kringle pastry

Two means of initiating blood clotting have been described, the *intrinsic pathway* and the *extrinsic pathway*. The intrinsic clotting pathway is activated by exposure of anionic surfaces upon rupture of the endothelial lining of the blood vessels. The extrinsic pathway, which appears to be most crucial in blood clotting, is initiated when trauma exposes *tissue factor* (TF), an integral membrane glycoprotein. Upon exposure to the blood, tissue factor binds to factor VII to activate factor X. Both the intrinsic and extrinsic pathways lead to the activation of factor X (a serine protease), which in turn converts *prothrombin* into *thrombin*, the key protease in clotting. Thrombin then amplifies the clotting process by activating enzymes and factors that lead to the generation of yet more thrombin, an example of positive feedback. Note that the active forms of the clotting factors are designated with a subscript "a," whereas factors that are activated by thrombin are designated with an asterisk.

Prothrombin requires a vitamin K-dependent modification for activation

Thrombin is synthesized as a zymogen called *prothrombin*. The inactive molecule comprises four major domains, with the serine protease domain at its carboxyl terminus (Figure 10.26). The first domain, called the *gla domain*, is rich in γ carboxyglutamate residues (abbreviation gla), and the second and third domains are called *kringle domains* (named after a Danish pastry that they resemble). Vitamin K is required for the synthesis of γ carboxyglutamate, a strong chelator of Ca^{2+}. These three domains work in concert to keep prothrombin in an inactive form. Moreover, because it is rich in γ carboxyglutamate, the gla domain is able to bind Ca^{2+} (Figure 10.27). What is the effect of this binding? The binding of Ca^{2+} by prothrombin anchors the zymogen to phospholipid membranes derived from blood platelets after injury. This binding is crucial because it brings prothrombin into close proximity to two clotting proteins, factor X_a and factor V_a (a stimulatory protein), that catalyze its conversion into thrombin. Activation is begun by proteolytic cleavage of the bond between arginine 274 and threonine 275 to release a fragment containing the first three domains. Cleavage of the bond between arginine 323 and isoleucine 324 yields active thrombin.

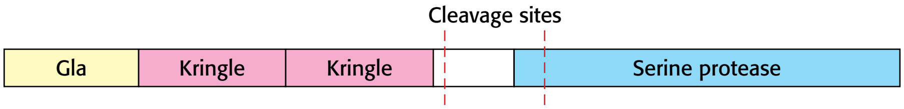


FIGURE 10.26 Modular structure of prothrombin. Cleavage of two peptide bonds yields thrombin. All the γ-carboxyglutamate residues are in the gla domain.

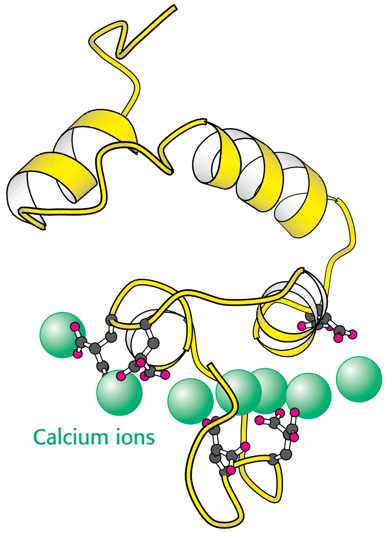


FIGURE 10.27 The calcium-binding region of prothrombin. Prothrombin binds calcium ions with the modified amino acid γ-carboxyglutamate (red). [Drawn from 2PF2.pdb.]

Fibrinogen is converted by thrombin into a fibrin clot

The best-characterized part of the clotting process is the final step in the cascade: the conversion of *fibrinogen* into fibrin by thrombin. Fibrinogen is made up of three globular units connected by two rods (Figure 10.28). This 340-kDa protein consists of six chains: two each of Aα, Bβ, and γ. The rod regions are triple-stranded α-helical coiled coils, a recurring motif in proteins (Section 2.3). Thrombin cleaves four *arginine–glycine peptide bonds* in the central globular region of fibrinogen (p. 217). On cleavage, an A peptide of 18 residues is released from each of the two Aα chains, as is a B peptide of 20 residues from each of the two Bβ chains. These A and B peptides are called *fibrinopeptides*. A fibrinogen molecule

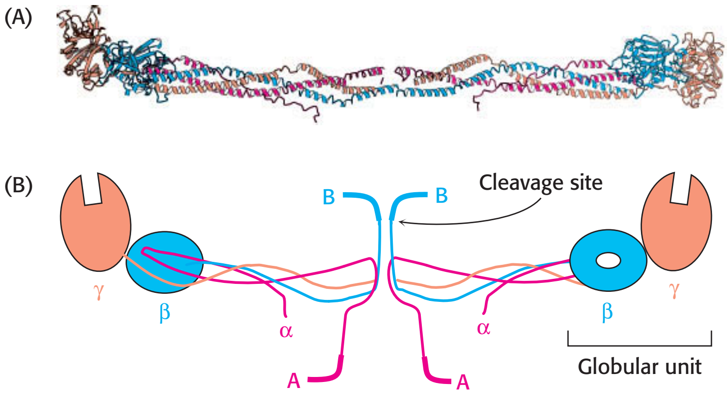


FIGURE 10.28 Structure of a fibrinogen molecule. (A) A ribbon diagram. The two rod regions are α-helical coiled coils, connected to a globular region at each end. The structure of the central globular region has not been determined. (B) A schematic representation showing the positions of the fibrinopeptides A and B. [Part A drawn from 1DEQ.pdb.]

devoid of these fibrinopeptides is called a *fibrin monomer* and has the subunit structure $(\alpha\beta\gamma)_2$.

Fibrin monomers spontaneously assemble into ordered fibrous arrays called *fibrin*. Electron micrographs and low-angle x-ray patterns show that fibrin has a periodic structure that repeats every 23 nm (Figure 10.29). Higher-resolution images reveal how the removal of the fibrinopeptides permits the fibrin monomers to come together to form fibrin. The homologous β and γ chains have globular domains at the carboxyl-terminal ends (Figure 10.30). These domains have binding "holes" that interact with peptides. The β domain is specific for sequences of the form H_3N^+-Gly-His-Arg-, whereas the γ domain binds H_3N^+-Gly-Pro-Arg-. Exactly these sequences (sometimes called "knobs") are exposed at the amino-terminal ends of the β and α chains, respectively, on thrombin cleavage. The knobs of the α subunits fit into the holes on the γ subunits of another monomer to form a protofibril. This protofibril is extended when the knobs of the β subunits fit into the holes of β subunits of other protofibrils. Thus, analogous to the activation of chymotrypsinogen, peptide-bond cleavage exposes new amino termini that can participate in specific interactions. The newly formed "soft clot" is stabilized by the formation of amide bonds between the side chains of lysine and glutamine residues in different monomers.

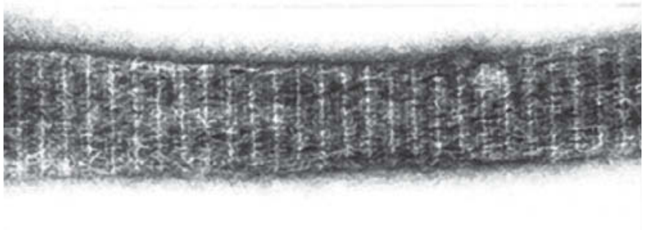


FIGURE 10.29 Electron micrograph of fibrin. The 23-nm period along the fiber axis is half the length of a fibrinogen molecule. [From John L. Woodhead et al., "The Ultrastructure of Fibrinogen Caracas II Molecules, Fibers, and Clots," *J. Biol. Chem.* 271(9):4946–4953, 1996, Mar 1. © American Society for Biochemistry and Molecular Biology.]

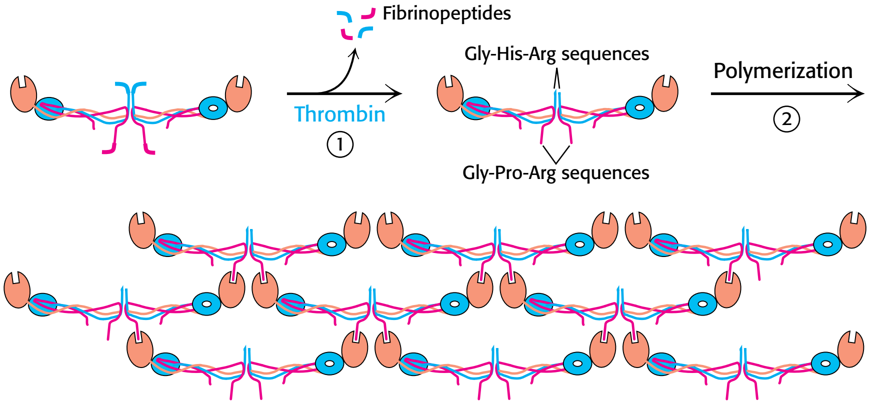
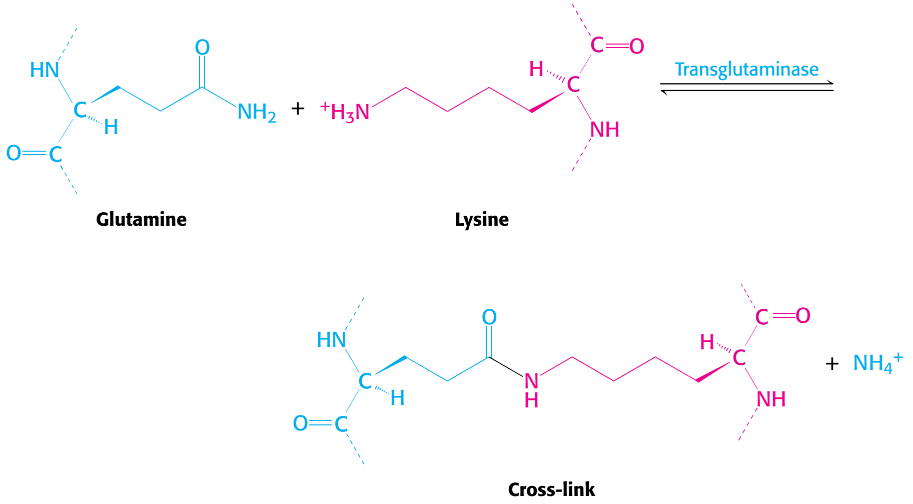


FIGURE 10.30 Formation of a fibrin clot. (1) Thrombin cleaves fibrinopeptides A and B from the central globule of fibrinogen. (2) Globular domains at the carboxyl-terminal ends of the β and γ chains interact with "knobs" exposed at the amino-terminal ends of the β and γ chains to form clots.


HN
O
NH$_2$
+
^+H_3N
C
H
O=C
H
C=O
C
NH
Transglutaminase
Glutamine
Lysine
HN
O
N
H
C=O
H
C
NH
+ NH_4^+
Cross-link

This cross-linking reaction is catalyzed by *transglutaminase (factor XIII$_a$)*, which itself is activated from the protransglutaminase form by thrombin.

Vitamin K is required for the formation of γ-carboxyglutamate

Vitamin K (Figure 10.31) has been known for many years to be essential for the synthesis of prothrombin and several other clotting factors. Indeed, it is called vitamin K because a deficiency in this vitamin results in defective blood *k*oagulation (Scandinavian spelling). After ingestion, vitamin K is reduced to a dihydro derivative that is required by γ-glutamyl carboxylase to convert the first 10 glutamate residues in the amino-terminal region of prothrombin into γ-carboxyglutamate (Figure 10.32).

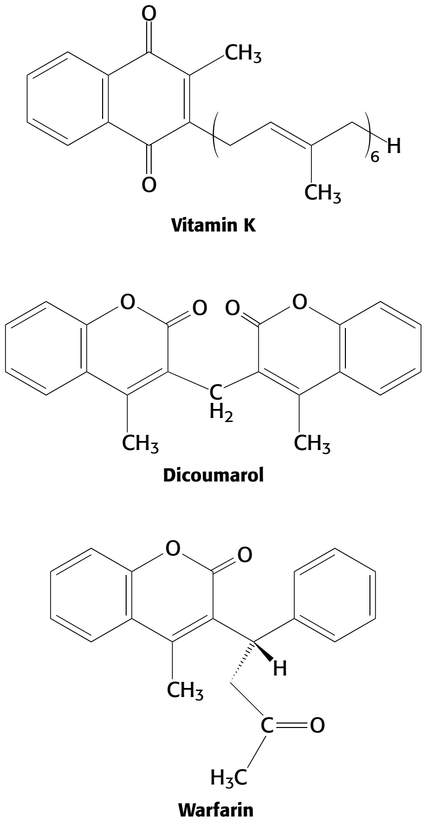

O
CH$_3$
O
6
H
CH$_3$
Vitamin K
O
O
O
O
C
H$_2$
CH$_3$
CH$_3$
Dicoumarol
O
O
H
CH$_3$
C=O
H$_3$C
Warfarin

FIGURE 10.31 Structures of vitamin K and two antagonists, dicoumarol and warfarin.

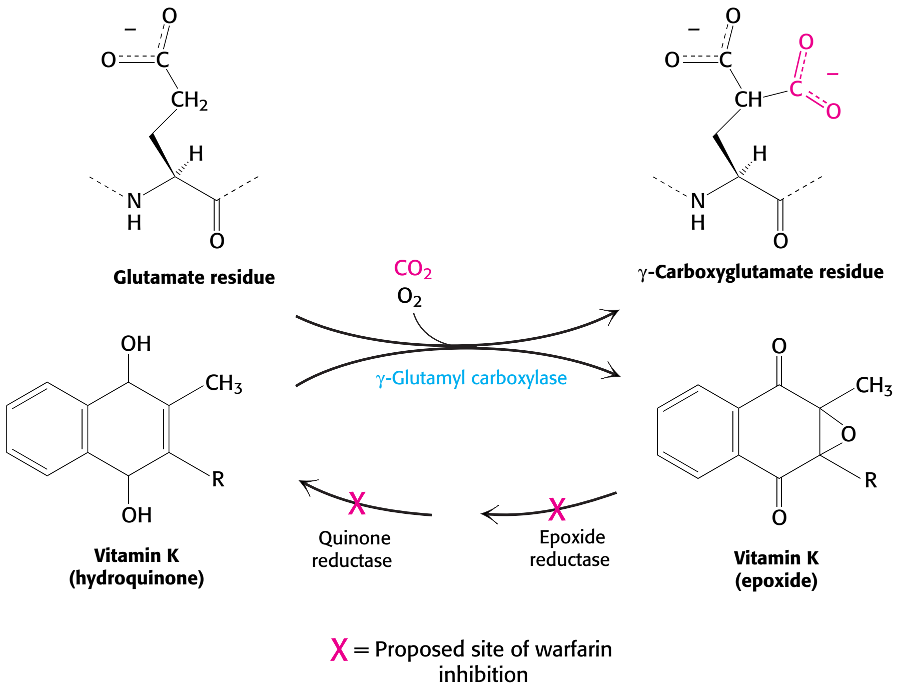

O
–
O
C
CH$_2$
H
N
H
O
Glutamate residue
O
–
O
C
O
–
CH
C
O
H
N
H
O
γ-Carboxyglutamate residue
CO$_2$
O$_2$
γ-Glutamyl carboxylase
OH
CH$_3$
R
OH
Vitamin K (hydroquinone)
Quinone reductase
Epoxide reductase
O
CH$_3$
O
R
O
Vitamin K (epoxide)
X = Proposed site of warfarin inhibition

FIGURE 10.32 Synthesis of γ-carboxyglutamate by γ-glutamyl carboxylase. The formation of γ-carboxyglutamate requires the hydroquinone derivative of vitamin K, which is regenerated from the epoxide derivative by the sequential action of epoxide reductase and quinone reductase, both of which are inhibited by warfarin.

Recall that γ-carboxyglutamate, a strong chelator of Ca^{2+}, is required for the activation of prothrombin (p. 304). *Dicoumarol*, which is found in spoiled sweet clover, causes a fatal hemorrhagic disease in cattle fed on this hay. Cows fed dicoumarol synthesize an abnormal prothrombin that does not bind Ca^{2+}, in contrast with normal prothrombin. Dicoumarol was the first *anticoagulant* used to prevent thromboses in patients prone to clot formation. However, it is seldom used now because of poor absorption and gastrointestinal side effects. *Warfarin*, another vitamin K antagonist, is commonly administered as an anticoagulant. Warfarin inhibits the keto reductase and quinone reductase that are required to regenerate the dihydro derivative of vitamin K (Figure 10.32). Dicoumarol, warfarin, and their chemical derivatives serve as effective rat poisons.

An account of a hemorrhagic disposition existing in certain families

"About seventy or eighty years ago, a woman by the name of Smith settled in the vicinity of Plymouth, New Hampshire, and transmitted the following idiosyncrasy to her descendants. It is one, she observed, to which her family is unfortunately subject and has been the source not only of great solicitude, but frequently the cause of death. If the least scratch is made on the skin of some of them, as mortal a hemorrhage will eventually ensue as if the largest wound is inflicted. . . . It is a surprising circumstance that the males only are subject to this strange affection, and that all of them are not liable to it. . . . Although the females are exempt, they are still capable of transmitting it to their male children."

John Otto (1803)

The clotting process must be precisely regulated

There is a fine line between hemorrhage and thrombosis, the formation of blood clots in blood vessels. Clots must form rapidly yet remain confined to the area of injury. What are the mechanisms that normally limit clot formation to the site of injury? The lability of clotting factors contributes significantly to the control of clotting. Activated factors are short-lived because they are diluted by blood flow, removed by the liver, and degraded by proteases. For example, the stimulatory protein factors V_a and $VIII_a$ are digested by protein C, a protease that is switched on by the action of thrombin. *Thus, thrombin has a dual function: it catalyzes the formation of fibrin and it initiates the deactivation of the clotting cascade.*

Specific inhibitors of clotting factors are also critical in the termination of clotting. For instance, *tissue factor pathway inhibitor* (TFPI) inhibits the complex of $TF–VII_a–X_a$ that activates thrombin. Another key inhibitor is *antithrombin III*, a member of the serpin family of protease inhibitors (p. 302) that forms an irreversible inhibitory complex with thrombin. Antithrombin III resembles α_1-antitrypsin except that it inhibits thrombin much more strongly than it inhibits elastase (Figure 10.23). Antithrombin III also blocks other serine proteases in the clotting cascade—namely, factors XII_a, XI_a, IX_a, and X_a. The inhibitory action of antithrombin III is enhanced by *heparin*, a negatively charged polysaccharide (Section 11.3) found in mast cells near the walls of blood vessels and on the surfaces of endothelial cells (Figure 10.33). Heparin acts as an *anticoagulant* by increasing the rate of formation of irreversible complexes between antithrombin III and the serine protease clotting factors.

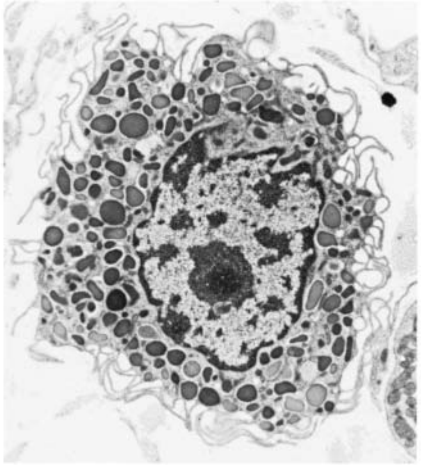


FIGURE 10.33 Electron micrograph of a mast cell. Heparin and other molecules in the dense granules are released into the extracellular space when the cell is triggered to secrete. [Courtesy of Lynne Mercer.]

The importance of the ratio of thrombin to antithrombin is illustrated in the case of a 14-year-old boy who died of a bleeding disorder because of a mutation in his α_1-antitrypsin, which normally inhibits elastase. Methionine 358 in α_1-antitrypsin's binding pocket for elastase was replaced by arginine, resulting in a change in specificity from an elastase inhibitor to a thrombin inhibitor. Activity of α_1-antitrypsin normally increases markedly after injury to counteract excess elastase arising from stimulated neutrophils. The mutant α_1-antitrypsin caused the patient's thrombin activity to drop to such a low level that hemorrhage ensued. We see here a striking example of how a change of a single residue in a protein can dramatically alter specificity and an example of the critical importance of having the right amount of a protease inhibitor.

Antithrombin limits the extent of clot formation, but what happens to the clots themselves? Clots are not permanent structures but are designed to

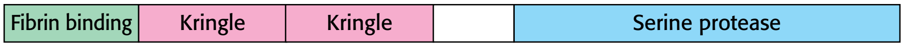


FIGURE 10.34 Modular structure of tissue-type plasminogen activator (TPA).

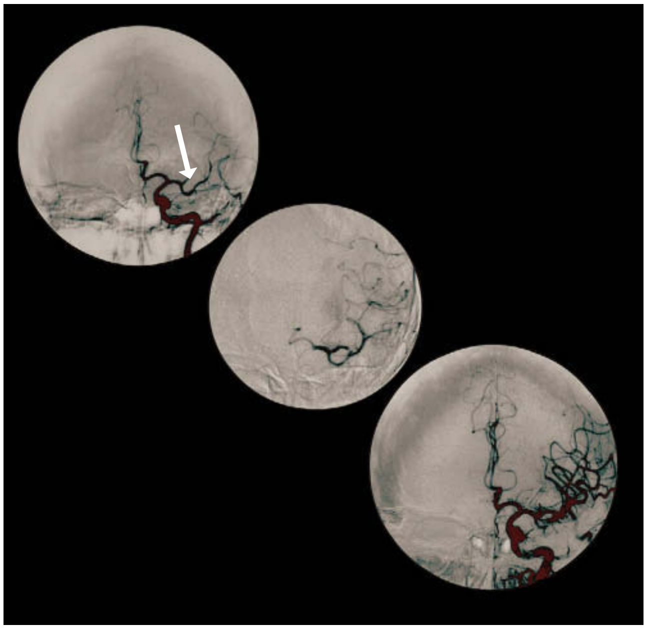


FIGURE 10.35 The effect of tissue-type plasminogen activator. Angiographic images demonstrate the effect of TPA administration. The top left image shows an occluded cerebral artery (arrow) prior to TPA injection. The middle image indicates the site of injection. The lower right image, made several hours after injection, reveals the restoration of blood flow to the cerebral artery. [Medical Body Scans/Science Source.]

dissolve when the structural integrity of damaged areas is restored. Fibrin is degraded by *plasmin,* a serine protease that hydrolyzes peptide bonds in the coiled-coil regions. Plasmin molecules can diffuse through aqueous channels in the porous fibrin clot to cut the accessible connector rods. Plasmin is formed by the proteolytic activation of *plasminogen,* an inactive precursor that has a high affinity for the fibrin clots. This conversion is carried out by *tissue-type plasminogen activator* (TPA), a 72-kDa protein that has a domain structure closely related to that of prothrombin (Figure 10.34). However, a domain that targets TPA to fibrin clots replaces the membrane-targeting gla domain of prothrombin. The TPA bound to fibrin clots swiftly activates adhering plasminogen. In contrast, TPA activates free plasminogen very slowly. The gene for TPA has been cloned and expressed in cultured mammalian cells. TPA administered at the onset of a heart attack or a stroke caused by a blood clot increases the likelihood of survival without physical or cognitive disabilities (Figure 10.35).

Hemophilia revealed an early step in clotting

Some important breakthroughs in the elucidation of clotting pathways have come from studies of patients with bleeding disorders. *Classic hemophilia,* or *hemophilia A,* is the best-known clotting defect. This disorder is genetically transmitted as a sex-linked recessive characteristic. *In classic hemophilia, factor VIII (antihemophilic factor) of the intrinsic pathway is missing or has markedly reduced activity.* Although factor VIII is not itself a protease, it markedly stimulates the activation of factor X, the final protease of the intrinsic pathway, by factor IX_a, a serine protease (Figure 10.36). Thus, activation of the intrinsic pathway is severely impaired in hemophilia.

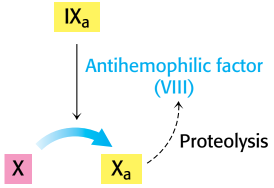


FIGURE 10.36 Action of antihemophilic factor. Antihemophilic factor (Factor VIII) stimulates the activation of factor X by factor IX_a. Interestingly, the activity of factor VIII is markedly increased by limited proteolysis by thrombin. This positive feedback amplifies the clotting signal and accelerates clot formation after a threshold has been reached.

In the past, hemophiliacs were treated with transfusions of a concentrated plasma fraction containing factor VIII. This therapy carried the risk of infection. Indeed, many hemophiliacs contracted hepatitis and, more recently, AIDS. A safer source of factor VIII was urgently needed. With the use of biochemical purification and recombinant DNA techniques, the gene for factor VIII was isolated and expressed in cells grown in culture. Recombinant factor VIII purified from these cells has largely replaced plasma concentrates in treating hemophilia.

SUMMARY

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway

Allosteric proteins constitute an important class of proteins whose biological activity can be regulated. Specific regulatory molecules can modulate the activity of allosteric proteins by binding to distinct regulatory sites, separate from the functional sites. These proteins have multiple functional sites, which display cooperativity as evidenced by a sigmoidal dependence of function on substrate concentration. Aspartate transcarbamoylase (ATCase), one of the best-understood allosteric enzymes, catalyzes the synthesis of *N*-carbamoylaspartate, the first intermediate in the synthesis of pyrimidines. ATCase is feedback inhibited by CTP, the final product of the pathway. ATP reverses this inhibition. ATCase consists of separable catalytic (c_3) subunits (which bind the substrates) and regulatory (r_2) subunits (which bind CTP and ATP). The inhibitory effect of CTP, the stimulatory action of ATP, and the cooperative binding of substrates are mediated by large changes in quaternary structure. On binding substrates, the c_3 subunits of the c_6r_6 enzyme move apart and reorient themselves. This allosteric transition is highly concerted. All subunits of an ATCase molecule simultaneously interconvert from the T (low-affinity) to the R (high-affinity) state.

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages

Isozymes differ in structural characteristics but catalyze the same reaction. They provide a means of fine-tuning metabolism to meet the needs of a given tissue or developmental stage. The results of gene-duplication events provide the means for subtle regulation of enzyme function.

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity

The covalent modification of proteins is a potent means of controlling the activity of enzymes and other proteins. Phosphorylation is a common type of reversible covalent modification. Signals can be highly amplified by phosphorylation because a single kinase can act on many target molecules. The regulatory actions of protein kinases are reversed by protein phosphatases, which catalyze the hydrolysis of attached phosphoryl groups.

Cyclic AMP serves as an intracellular messenger in the transduction of many hormonal and sensory stimuli. Cyclic AMP switches on protein kinase A, a major multifunctional kinase, by binding to the regulatory subunit of the enzyme, thereby releasing the active catalytic subunits of PKA. In the absence of cAMP, the catalytic sites of PKA are occupied by pseudosubstrate sequences of the regulatory subunit.

10.4 Many Enzymes Are Activated by Specific Proteolytic Cleavage

The activation of an enzyme by the proteolytic cleavage of one or a few peptide bonds is a recurring control mechanism seen in processes as diverse as the activation of digestive enzymes and blood clotting. The inactive precursor is a zymogen (proenzyme). Trypsinogen is activated by enteropeptidase or trypsin, and trypsin then activates a host of other zymogens, leading to the digestion of foodstuffs. For instance, trypsin converts chymotrypsinogen, a zymogen, into active chymotrypsin by hydrolyzing a single peptide bond.

A striking feature of the clotting process is that it is accomplished by a cascade of zymogen conversions, in which the activated form of one clotting factor catalyzes the activation of the next precursor. Many of the activated clotting factors are serine proteases. In the final step of clot formation, fibrinogen, a highly soluble molecule in the plasma, is converted by thrombin into fibrin by the hydrolysis of four arginine–glycine bonds. The resulting fibrin monomer spontaneously forms long, insoluble fibers called fibrin. Zymogen activation is also essential in the lysis of clots. Plasminogen is converted into plasmin, a serine protease that cleaves fibrin, by tissue-type plasminogen activator. Although zymogen activation is irreversible, specific inhibitors of some proteases exert control. The irreversible protein inhibitor antithrombin III holds blood clotting in check in the clotting cascade.

KEY TERMS

cooperativity (p. 285)
feedback (end-product) inhibition (p. 287)
allosteric (regulatory) site (p. 287)
homotropic effect (p. 290)
concerted model (p. 290)
sequential model (p. 290)
heterotropic effect (p. 292)
isozyme (isoenzyme) (isoform) (p. 292)
covalent modification (p. 293)
protein kinase (p. 294)
consensus sequence (p. 295)
protein phosphatase (p. 295)
protein kinase A (PKA) (p. 297)
pseudosubstrate sequence (p. 297)
zymogen (proenzyme) (p. 299)
enzymatic cascade (p. 303)
hemostasis (p. 303)
intrinsic pathway (p. 304)
extrinsic pathway (p. 304)

PROBLEMS

1. *Context please.* The allosteric properties of aspartate transcarbamoylase have been discussed in detail in this chapter. What is the function of aspartate transcarbamoylase?

2. *Activity profile.* A histidine residue in the active site of aspartate transcarbamoylase is thought to be important in stabilizing the transition state of the bound substrates. Predict the pH dependence of the catalytic rate, assuming that this interaction is essential and dominates the pH-activity profile of the enzyme. (See equations on p. 16.)

3. *Knowing when to say when.* What is feedback inhibition? Why is it a useful property?

4. *Knowing when to get going.* What is the biochemical rationale for ATP serving as a positive regulator of ATCase?

5. *No T.* What would be the effect of a mutation in an allosteric enzyme that resulted in a T/R ratio of 0?

6. *Turned upside down.* An allosteric enzyme that follows the concerted model has a T/R ratio of 300 in the absence of substrate. Suppose that a mutation reversed the ratio. How would this mutation affect the relation between the rate of the reaction and the substrate concentration?

7. *Partners.* As shown in Figure 10.2, CTP inhibits ATCase; however, the inhibition is not complete. Can you suggest another molecule that might enhance the inhibition of ATCase? Hint: See Figure 25.2.

8. *RT equilibrium.* Differentiate between homotropic and heterotropic effectors.

9. *Restoration project.* If isolated regulatory subunits and catalytic subunits of ATCase are mixed, the native enzyme is reconstituted. What is the biological significance of the observation?

10. *Because it's an enzyme.* X-ray crystallographic studies of ATCase in the R form required the use of the bisubstrate analog PALA. Why was this analog, a competitive inhibitor, used instead of the actual substrates?

11. *Allosteric switching.* A substrate binds 100 times as tightly to the R state of an allosteric enzyme as to its T state. Assume that the concerted (MWC) model applies to this enzyme. (See equations for the Concerted Model in the Appendix to Chapter 7.)

(a) By what factor does the binding of one substrate molecule per enzyme molecule alter the ratio of the concentrations of enzyme molecules in the R and T states?

(b) Suppose that L, the ratio of [T] to [R] in the absence of substrate, is 10^7 and that the enzyme contains four binding sites for substrate. What is the ratio of enzyme molecules in the R state to those in the T state in the presence of saturating amounts of substrate, assuming that the concerted model is obeyed?

12. *Allosteric transition.* Consider an allosteric protein that obeys the concerted model. Suppose that the ratio of T to R formed in the absence of ligand is 10^5, $K_T = 2$ mM, and $K_R = 5\ \mu$M. The protein contains four binding sites for ligand. What is the fraction of molecules in the R form when 0, 1, 2, 3, and 4 ligands are bound? (See equations for the Concerted Model in the Appendix to Chapter 7.)

13. *Negative cooperativity.* You have isolated a dimeric enzyme that contains two identical active sites. The binding of substrate to one active site decreases the substrate affinity of the other active site. Can the concerted model account for this negative cooperativity? Hint: See Section 7.2.

14. *A new view of cooperativity.* Draw a double-reciprocal plot for a typical Michaelis–Menten enzyme and an allosteric enzyme that have the same V_{max} and K_M. Draw a double reciprocal plot for the same allosteric enzyme in the presence of an allosteric inhibitor and an allosteric stimulator.

15. *Paradoxical at first glance.* Recall that phosphonacetyl-L-aspartate (PALA) is a potent inhibitor of ATCase because

it mimics the two physiological substrates. However, in the presence of substrates, low concentrations of this unreactive bisubstrate analog *increase* the reaction velocity. On the addition of PALA, the reaction rate increases until an average of three molecules of PALA are bound per molecule of enzyme. This maximal velocity is 17-fold greater than it is in the absence of PALA. The reaction rate then decreases to nearly zero on the addition of three more molecules of PALA per molecule of enzyme. Why do low concentrations of PALA activate ATCase?

16. *Regulation energetics.* The phosphorylation and dephosphorylation of proteins is a vital means of regulation. Protein kinases attach phosphoryl groups, whereas only a phosphatase will remove the phosphoryl group from the target protein. What is the energy cost of this means of covalent regulation?

17. *Vive la différence*. What is an isozyme?

18. *Fine-tuning biochemistry*. What is the advantage for an organism to have isozymic forms of an enzyme?

19. *Making matches*.

(a) ATCase _________	1. Protein phosphorylation catalyst
(b) T state _________	2. Required to modify glutamate
(c) R state _________	3. Activates a particular kinase
(d) Phosphorylation _________	4. Proenzyme
(e) Kinase _________	5. Activates trypsin
(f) Phosphatase _________	6. Common covalent modification
(g) cAMP _________	7. Inhibited by CTP
(h) Zymogen _________	8. Less-active state of an allosteric protein
(i) Enteropeptidase _________	9. Initiates extrinsic pathway
(j) Vitamin K _________	10. Forms fibrin
(k) Thrombin _________	11. More-active state of an allosteric protein
(l) Tissue factor _________	12. Removes phosphates

20. *Powering change*. Phosphorylation is a common covalent modification of proteins in all forms of life. What energetic advantages accrue from the use of ATP as the phosphoryl donor?

21. *No going back*. What is the key difference between regulation by covalent modification and specific proteolytic cleavage?

22. *Zymogen activation.* When very low concentrations of pepsinogen are added to acidic media, how does the half-time for activation depend on zymogen concentration?

23. *No protein shakes advised*. Predict the physiological effects of a mutation that resulted in a deficiency of enteropeptidase.

24. *A revealing assay.* Suppose that you have just examined a young boy with a bleeding disorder highly suggestive of classic hemophilia (factor VIII deficiency). Because of the late hour, the laboratory that carries out specialized coagulation assays is closed. However, you happen to have a sample of blood from a classic hemophiliac whom you admitted to the hospital an hour earlier. What is the simplest and most rapid test that you can perform to determine whether your present patient also is deficient in factor VIII activity?

25. *Counterpoint*. The synthesis of factor X, like that of prothrombin, requires vitamin K. Factor X also contains γ-carboxyglutamate residues in its amino-terminal region. However, activated factor X, in contrast with thrombin, retains this region of the molecule. What is a likely functional consequence of this difference between the two activated species?

26. *A discerning inhibitor.* Antithrombin III forms an irreversible complex with thrombin but not with prothrombin. What is the most likely reason for this difference in reactivity?

27. *Drug design*. A drug company has decided to use recombinant DNA methods to prepare a modified α_1-antitrypsin that will be more resistant to oxidation than is the naturally occurring inhibitor. Which single amino acid substitution would you recommend?

28. *Blood must flow*. Why is inappropriate blood-clot formation dangerous?

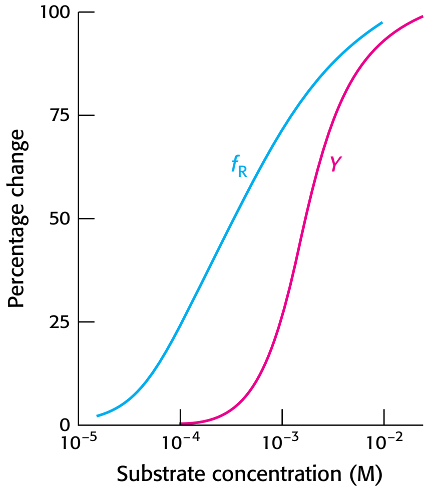
29. *Hemostasis.* Thrombin functions in both coagulation and fibrinolysis. Explain.

30. *Dissolution row*. What is tissue-type plasminogen activator and what is its role in preventing heart attacks?

31. *Joining together*. What differentiates a soft clot from a mature clot?

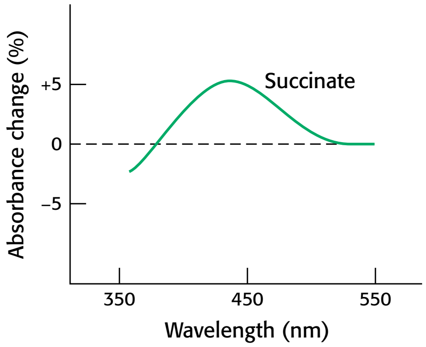
Data Interpretation Problems

32. *Distinguishing between models.* The following graph shows the fraction of an allosteric enzyme in the R state (f_R) and the fraction of active sites bound to substrate (Y) as a function of substrate concentration. Which model, the concerted or sequential, best explains these results?



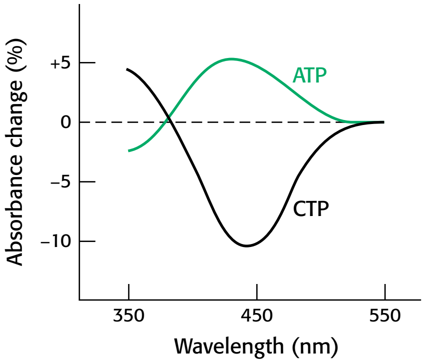
[Data from M. W. Kirschner and H. K. Schachman. *Biochemistry* 12:2997–3004, 1966.]

33. *Reporting live from ATCase 1.* ATCase underwent reaction with tetranitromethane to form a colored nitrotyrosine group ($\lambda_{\max}$ = 430 nm) in each of its catalytic chains. The absorption by this reporter group depends on its immediate environment. An essential lysine residue at each catalytic site also was modified to block the binding of substrate. Catalytic trimers from this doubly modified enzyme were then combined with native trimers to form a hybrid enzyme. The absorption by the nitrotyrosine group was measured on addition of the substrate analog succinate. What is the significance of the alteration in the absorbance at 430 nm?



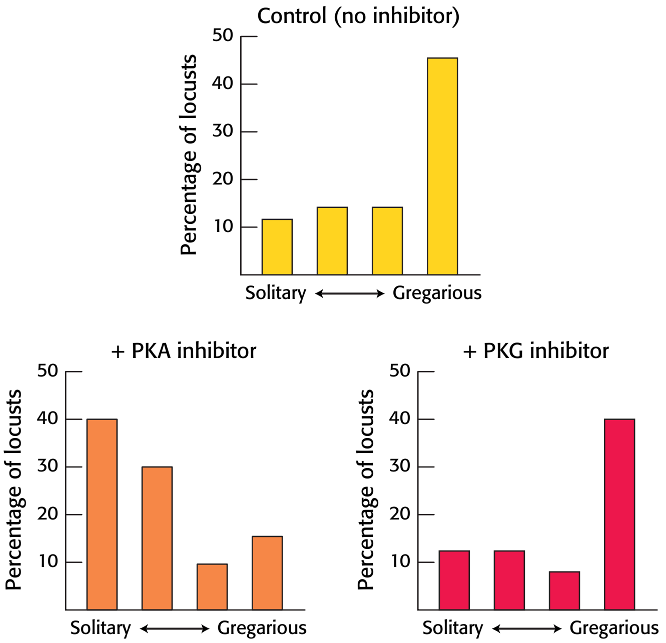
[Data from H. K. Schachman. *J. Biol. Chem*. 263:18583–18586, 1988.]

34. *Reporting live from ATCase 2.* A different ATCase hybrid was constructed to test the effects of allosteric activators and inhibitors. Normal regulatory subunits were combined with nitrotyrosine-containing catalytic subunits. The addition of ATP in the absence of substrate increased the absorbance at 430 nm, the same change elicited by the addition of succinate (see the graph in Problem 33). Conversely, CTP in the absence of substrate decreased the absorbance at 430 nm. What is the significance of the changes in absorption of the reporter groups?



[Data from H. K. Schachman. *J. Biol. Chem*. 263:18583–18586, 1988.]

35. *Conviviality and PKA.* Recent studies have suggested that protein kinase A may be important in establishing behaviors in many organisms, including humans. One study investigated the role of PKA in locust behavior. Certain species of locust live solitary lives until crowded, at which point they become gregarious—they prefer the crowded life.



[Data from S. R. Ott, et al., *Proc. Natl. Acad. Sci. U. S. A.* 109(7):E381–7, 2012.]

Locusts were grouped together for one hour, and then allowed to stay with the group or move away. Prior to crowding, some insects were injected with a PKA inhibitor, a cyclic GMP-dependent kinase inhibitor or no inhibitor, as indicated. The results are shown above.

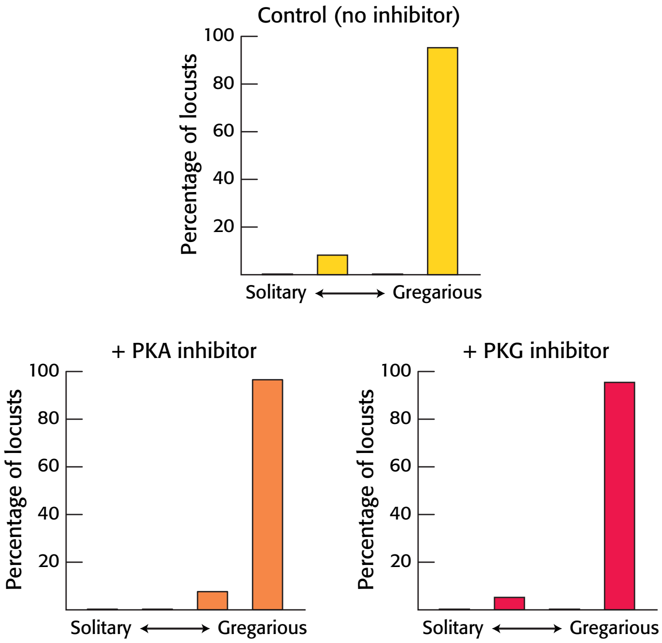
(a) What is the response of the control group to crowding?

(b) What is the result if the insects are first treated with PKA inhibitor? PKG inhibitor?

(c) What was the purpose of the experiment with the PKG inhibitor?

(d) What do these results suggest about the role of PKA in the transition from a solitary to a gregarious life style?

The above experiments were repeated on a different species of insect that is always gregarious. The results are shown below.



[Data from S. R. Ott, et al., *Proc. Natl. Acad. Sci. U. S. A.* 109(7):E381–7, 2012.]

(e) What do these results suggest about the role of PKA in insects that are always gregarious?

Chapter Integration Problems

36. *Repeating heptads.* Each of the three types of fibrin chains contains repeating heptapeptide units (*abcdefg*) in which residues *a* and *d* are hydrophobic. Propose a reason for this regularity.

37. *Density matters.* The sedimentation value of aspartate transcarbamoylase decreases when the enzyme switches to the R state. On the basis of the allosteric properties of the enzyme, explain why the sedimentation value decreases.

38. *Too tight a grip.* Trypsin cleaves proteins on the carboxyl side of lysine. Trypsin inhibitor has a lysine residue, and binds to trypsin, yet it is not a substrate. Explain.

39. *Apparently not the four-leaf variety.* Cows that graze on spoiled sweet clover, which contains dicoumarol, die from hemorrhagic disease. The cause of death is defective prothrombin. However, the amino acid composition of the defective prothrombin is identical to that of normal prothrombin. What is the mechanism of action of dicoumarol? Why are the amino acid compositions of the defective and normal prothrombin the same?

Mechanism Problems

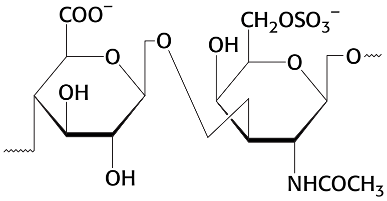
40. *Aspartate transcarbamoylase.* Write the mechanism (in detail) for the conversion of aspartate and carbamoyl phosphate into *N*-carbamoylaspartate. Include a role for the histidine residue present in the active site.

41. *Protein kinases.* Write a mechanism (in detail) for phosphorylation of a serine residue by ATP catalyzed by protein kinase. What groups might you expect to find in the enzyme's active site?

CHAPTER

11

Carbohydrates



Carbohydrates are important fuel molecules, but they play many other biochemical roles, including protection against high-impact forces. The cartilage of a runner's foot cushions the impact of each step she takes. A key component of cartilage are molecules called glycosaminoglycans, large polymers made up of many repeats of dimers such as the pair shown at the right. [Untitled x-ray/Nick Veasey/Getty Images.]

OUTLINE

11.1 Monosaccharides Are the Simplest Carbohydrates

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins

11.4 Lectins Are Specific Carbohydrate-Binding Proteins

For years, the study of carbohydrates was considered less exciting than many if not most topics of biochemistry. Carbohydrates were recognized as important fuels and structural components but were thought to be peripheral to most key activities of the cell. In essence, they were considered the underlying girders and fuel for a magnificent piece of biochemical architecture. This view has changed dramatically in the past few years. We have learned that cells of all organisms are coated in a dense and complex coat of carbohydrates. Secreted proteins are often extensively decorated with carbohydrates essential to a protein's function. The extracellular matrix in higher eukaryotes—the environment in which the cells live—is rich in secreted carbohydrates central to cell survival and cell-to-cell communication. Carbohydrates, carbohydrate-containing proteins, and specific carbohydrate-binding proteins are required for interactions that allow cells to form tissues, are the basis of human blood groups, and are used by a variety of pathogens to gain access to their hosts. Indeed, rather than being mere infrastructure components, carbohydrates supply details and enhancements to the biochemical architecture of the cell, helping to define the functionality and uniqueness of the cell.

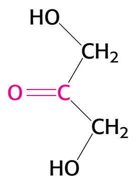
A key property of carbohydrates that allows their many functions is the tremendous structural *diversity* possible within this class of molecules. Carbohydrates are built from monosaccharides, which are small molecules—typically containing from three to nine carbon atoms that are bound to hydroxyl groups—that vary in size and in the stereochemical configuration at one or more carbon centers. These monosaccharides can be linked together to form a large variety of oligosaccharide structures. The sheer number of possible oligosaccharides makes this class of molecules information rich. This information, when attached to proteins, can augment the already immense diversity of proteins.

The realization of the importance of carbohydrates to so many aspects of biochemistry has spawned a field of study called *glycobiology*. Glycobiology is the study of the synthesis and structure of carbohydrates and how carbohydrates are attached to and recognized by other molecules such as proteins. Along with a new field comes a new "omics" to join genomics and proteomics—*glycomics*. Glycomics is the study of the glycome, all of the carbohydrates and carbohydrate-associated molecules that cells produce. Like the proteome, the glycome is dynamic, depending on cellular and environmental conditions. Unraveling oligosaccharide structures and elucidating the effects of their attachment to other molecules constitutes a tremendous challenge in the field of biochemistry.

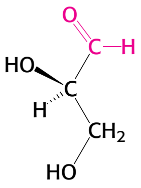
11.1 Monosaccharides Are the Simplest Carbohydrates

Carbohydrates are carbon-based molecules that are rich in hydroxyl groups. Indeed, the empirical formula for many carbohydrates is $(CH_2O)_n$—literally, a carbon hydrate. Simple carbohydrates are called *monosaccharides.* These simple sugars serve not only as fuel molecules but also as fundamental constituents of living systems. For instance, DNA has a backbone consisting of alternating phosphoryl groups and deoxyribose, a cyclic five-carbon sugar.

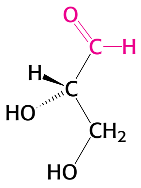
Monosaccharides are aldehydes or ketones that have two or more hydroxyl groups. The smallest monosaccharides, composed of three carbon atoms, are dihydroxyacetone and D- and L-glyceraldehyde.



Dihydroxyacetone
(a ketose)



D-Glyceraldehyde
(an aldose)



L-Glyceraldehyde
(an aldose)

Dihydroxyacetone is called a *ketose* because it contains a keto group (in red above), whereas glyceraldehyde is called an *aldose* because it contains an aldehyde group (also in red). They are referred to as *trioses* (tri- for three, referring to the three carbon atoms that they contain). Similarly, simple monosaccharides with four, five, six, and seven carbon atoms are called *tetroses, pentoses, hexoses,* and *heptoses,* respectively. Perhaps the monosaccharides of which we are most aware are the hexoses, such as glucose and

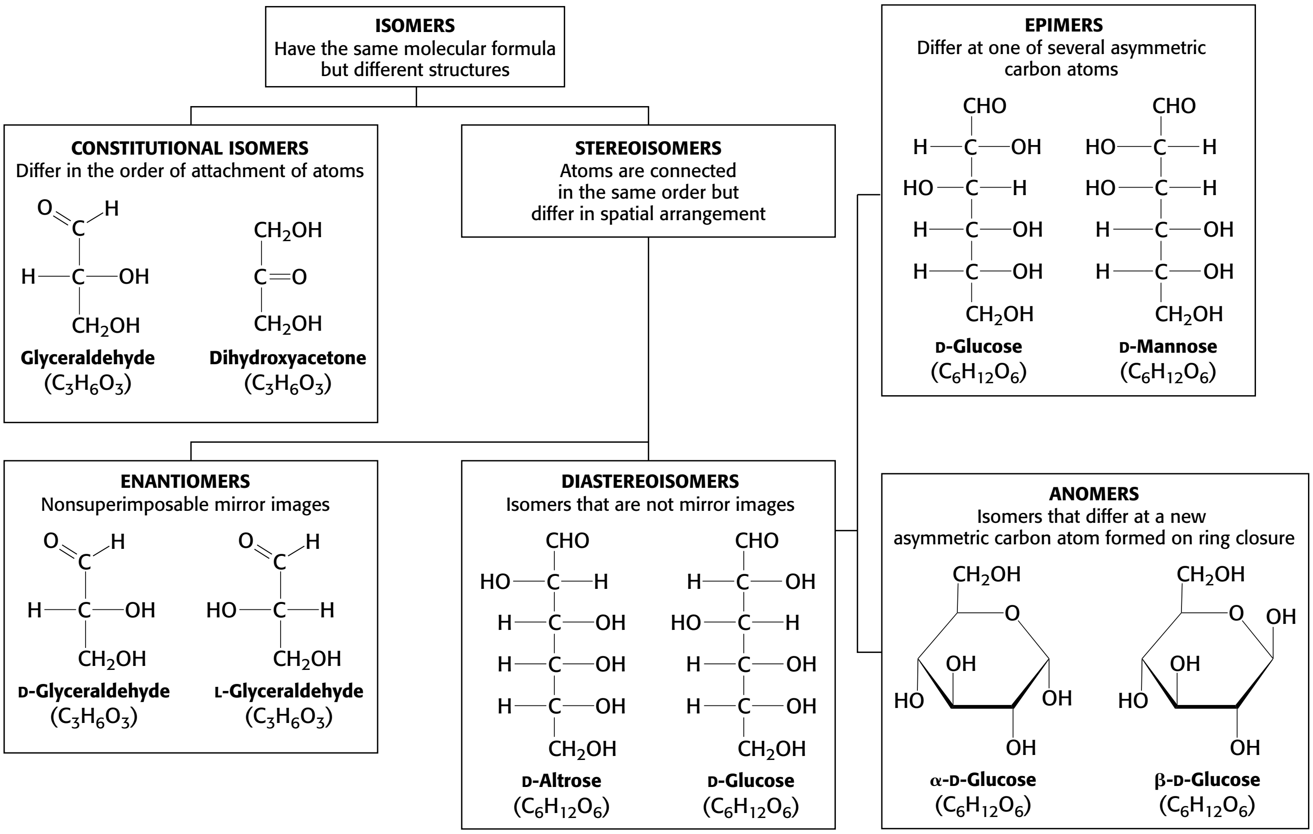

ISOMERS
Have the same molecular formula but different structures
CONSTITUTIONAL ISOMERS
Differ in the order of attachment of atoms
Glyceraldehyde
($C_3H_6O_3$)
Dihydroxyacetone
($C_3H_6O_3$)
STEREOISOMERS
Atoms are connected in the same order but differ in spatial arrangement
EPIMERS
Differ at one of several asymmetric carbon atoms
D-Glucose
($C_6H_{12}O_6$)
D-Mannose
($C_6H_{12}O_6$)
ENANTIOMERS
Nonsuperimposable mirror images
D-Glyceraldehyde
($C_3H_6O_3$)
L-Glyceraldehyde
($C_3H_6O_3$)
DIASTEREOISOMERS
Isomers that are not mirror images
D-Altrose
($C_6H_{12}O_6$)
D-Glucose
($C_6H_{12}O_6$)
ANOMERS
Isomers that differ at a new asymmetric carbon atom formed on ring closure
α-D-Glucose
($C_6H_{12}O_6$)
β-D-Glucose
($C_6H_{12}O_6$)

FIGURE 11.1 Isomeric forms of carbohydrates.

fructose. Glucose is an essential energy source for virtually all forms of life. Fructose is commonly used as a sweetener that is converted into glucose derivatives inside the cell.

Carbohydrates can exist in a dazzling variety of isomeric forms (Figure 11.1). Dihydroxyacetone and glyceraldehyde are *constitutional isomers* because they have identical molecular formulas but differ in how the atoms are ordered. *Stereoisomers* are isomers that differ in spatial arrangement. Recall from the discussion of amino acids (p. 29) that stereoisomers are designated as having either D or L configuration. Glyceraldehyde has a single asymmetric carbon atom and, thus, there are two stereoisomers of this sugar: D-glyceraldehyde and L-glyceraldehyde. These molecules are a type of stereoisomer called *enantiomers,* which are mirror images of each other. Most vertebrate monosaccharides have the D configuration. According to convention, the D and L isomers are determined by the configuration of the asymmetric carbon atom farthest from the aldehyde or keto group. Dihydroxyacetone is the only monosaccharide without at least one asymmetric carbon atom.

Monosaccharides made up of more than three carbon atoms have multiple asymmetric carbons, and so they can exist not only as enantiomers but also as *diastereoisomers,* isomers that are not mirror images of each other. The number of possible stereoisomers equals 2^n, where n is the number of asymmetric carbon atoms. Thus, a six-carbon aldose with 4 asymmetric carbon atoms can exist as 16 possible diastereoisomers, of which glucose is one such isomer.

Figure 11.2 shows the common sugars that we will see most frequently in our study of biochemistry. D-Ribose, the carbohydrate component of RNA, is a five-carbon aldose, as is deoxyribose, the monosaccharide component of deoxynucleotides. D-Glucose, D-mannose, and D-galactose are

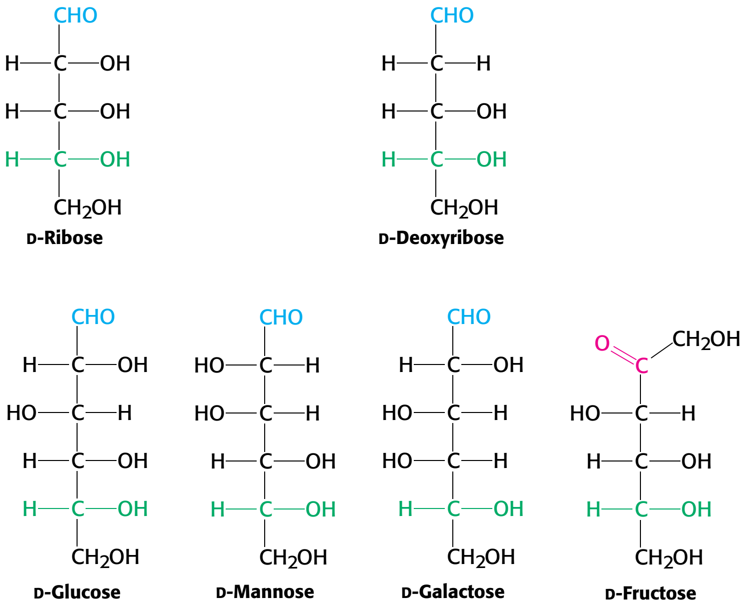


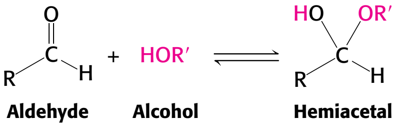
FIGURE 11.2 Common monosaccharides. Aldoses contain an aldehyde (shown in blue), whereas ketoses, such as fructose, contain a ketose (shown in red). The asymmetric carbon atom farthest from the aldehyde or ketone (shown in green) designates the structures as being in the D configuration.

abundant six-carbon aldoses. Note that D-glucose and D-mannose differ in configuration only at C-2, the carbon atom in the second position. Sugars that are diastereoisomers differing in configuration at only a single asymmetric center are *epimers*. Thus, D-glucose and D-mannose are epimeric at C-2; D-glucose and D-galactose are epimeric at C-4.

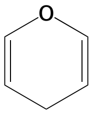
Note that ketoses have one less asymmetric center than aldoses with the same number of carbon atoms. D-Fructose is the most abundant ketohexose.

Many common sugars exist in cyclic forms

The predominant forms of ribose, glucose, fructose, and many other sugars in solution, as is the case inside the cell, are not open chains. Rather, the open-chain forms of these sugars cyclize into rings. The chemical basis for ring formation is that an aldehyde can react with an alcohol to form a *hemiacetal*.

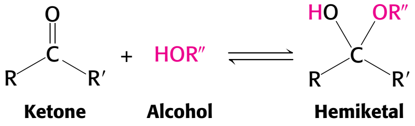


For an aldohexose such as glucose, a single molecule provides both the aldehyde and the alcohol: the C-1 aldehyde in the open-chain form of glucose reacts with the C-5 hydroxyl group to form an *intramolecular hemiacetal* (Figure 11.3). The resulting cyclic hemiacetal, a six-membered ring, is called *pyranose* because of its similarity to *pyran*.

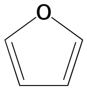


Pyran

Similarly, a ketone can react with an alcohol to form a *hemiketal*.



The C-2 keto group in the open-chain form of a ketohexose, such as fructose, can form an *intramolecular hemiketal* by reacting with either the



Furan

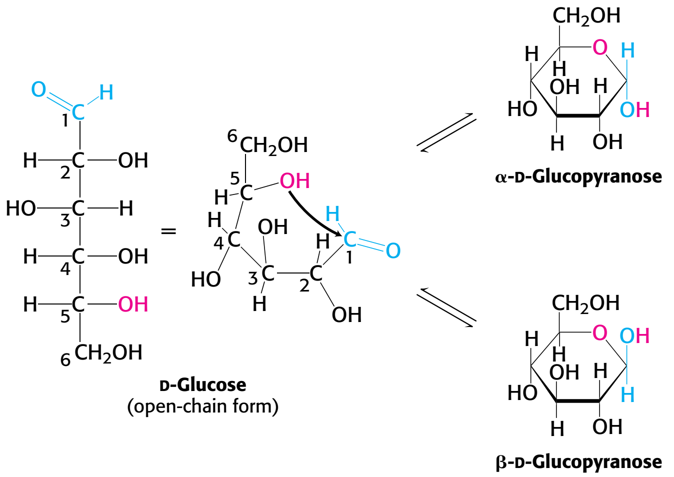


FIGURE 11.3 Pyranose formation. The open-chain form of glucose cyclizes when the C-5 hydroxyl group attacks the oxygen atom of the C-1 aldehyde group to form an intramolecular hemiacetal. Two anomeric forms, designated α and β, can result.

C-6 hydroxyl group to form a six-membered cyclic hemiketal or the C-5 hydroxyl group to form a five-membered cyclic hemiketal (Figure 11.4). The five-membered ring is called a *furanose* because of its similarity to *furan*.

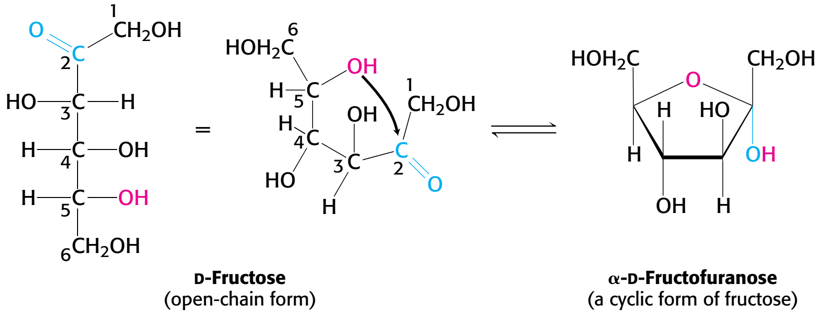
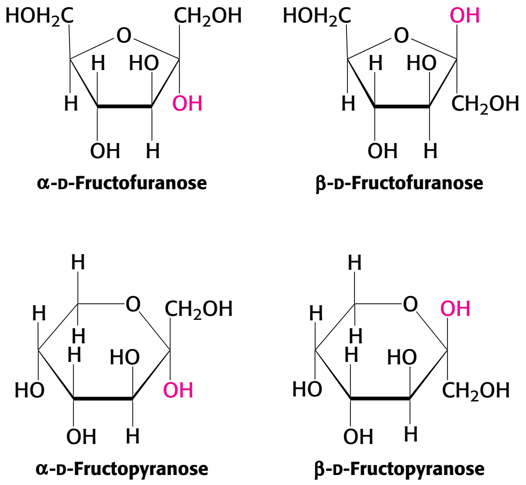


FIGURE 11.4 Furanose formation. The open-chain form of fructose cyclizes to a five-membered ring when the C-5 hydroxyl group attacks the C-2 ketone to form an intramolecular hemiketal. Two anomers are possible, but only the α anomer is shown.

The depictions of glucopyranose (glucose) and fructofuranose (fructose) shown in Figures 11.3 and 11.4 are *Haworth projections.* In such projections, the carbon atoms in the ring are not written out. The approximate plane of the ring is perpendicular to the plane of the paper, with the heavy line on the ring projecting toward the reader.

We have seen that carbohydrates can contain many asymmetric carbon atoms. An additional asymmetric center is created when a cyclic hemiacetal is formed, generating yet another diastereoisomeric form of sugars called *anomers*. In glucose, C-1 (the carbonyl carbon atom in the open-chain form) becomes an asymmetric center. Thus, two ring structures can be formed: α-D-glucopyranose and β-D-glucopyranose (Figure 11.3). For D sugars drawn as Haworth projections in the standard orientation as shown in Figure 11.3, the *designation α means that the hydroxyl group attached to C-1 is on the opposite side of the ring as C-6; β means that the hydroxyl group is on the same side of the ring as C-6.* The C-1 carbon atom is called the *anomeric carbon atom.* An equilibrium mixture of glucose contains approximately one-third α anomer, two-thirds β anomer, and <1% of the open-chain form.

The furanose-ring form of fructose also has anomeric forms, in which α and β refer to the hydroxyl groups attached to C-2, the anomeric carbon atom (Figure 11.4). Fructose forms both pyranose and furanose rings. The

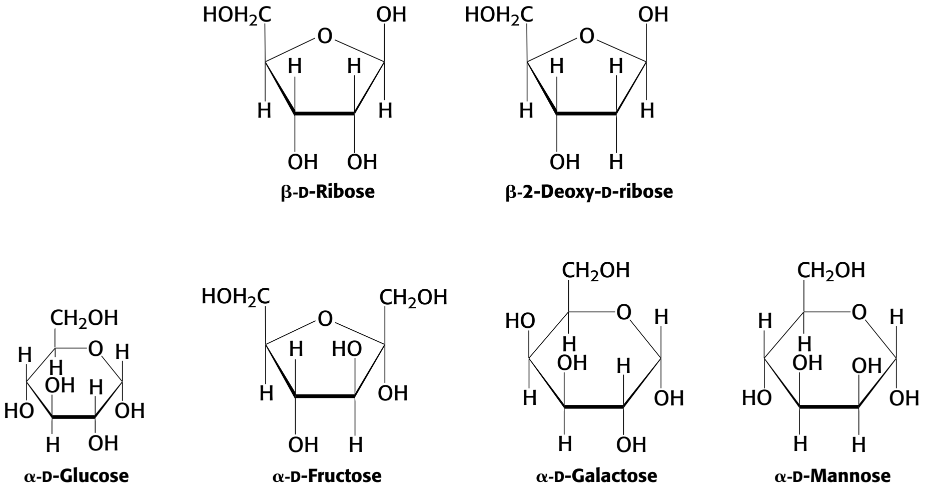

α-D-Fructofuranose **β-D-Fructofuranose**

α-D-Fructopyranose **β-D-Fructopyranose**

FIGURE 11.5 Ring structures of fructose. Fructose can form both five-membered furanose (top) and six-membered pyranose (bottom) rings. In each case, both α and β anomers are possible.

pyranose form predominates in fructose that is free in solution, and the furanose form predominates in many fructose derivatives (Figure 11.5).

β-D-Fructopyranose, found in honey, is one of the sweetest chemicals known. The β-D-fructofuranose form is not nearly as sweet. Heating converts β-fructopyranose into the β-fructofuranose form, reducing the sweetness of the solution. For this reason, corn syrup with a high concentration of fructose in the β-D-pyranose form is used as a sweetener in cold, but not hot, drinks. Figure 11.6 shows the common sugars discussed previously in their ring forms.

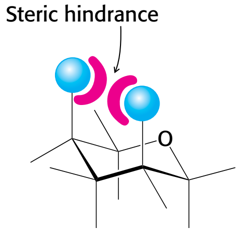

β-D-Ribose **β-2-Deoxy-D-ribose**

α-D-Glucose **α-D-Fructose** **α-D-Galactose** **α-D-Mannose**

FIGURE 11.6 Common monosaccharides in their ring forms.

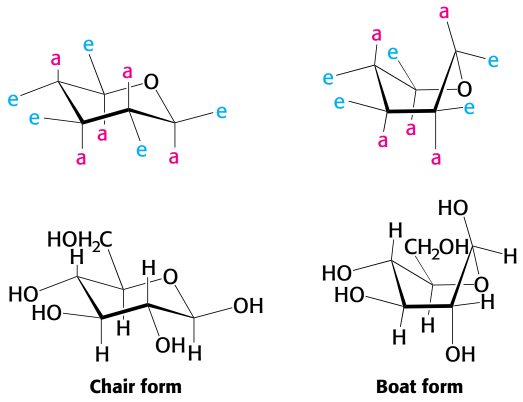
Pyranose and furanose rings can assume different conformations

The six-membered pyranose ring is not planar because of the tetrahedral geometry of its saturated carbon atoms. Instead, pyranose rings adopt two classes of conformations, termed chair and boat because of the resemblance to these objects (Figure 11.7). In the chair form, the substituents on the ring carbon atoms have two orientations: axial and equatorial. *Axial* bonds are nearly perpendicular to the average plane of the ring, whereas *equatorial* bonds are nearly parallel to this plane. Axial substituents sterically hinder each other if they emerge on the same side of the ring (e.g., 1,3-diaxial groups). In contrast, equatorial substituents are less



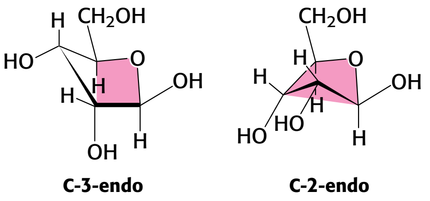
crowded. The chair form of *β-D-glucopyranose predominates because all axial positions are occupied by hydrogen atoms.* The bulkier —OH and —CH_2OH groups emerge at the less-hindered periphery. The boat form of glucose is disfavored because it is quite sterically hindered.

Furanose rings, like pyranose rings, are not planar. They can be puckered so that four atoms are nearly coplanar and the fifth is about 0.5 Å away from this plane (Figure 11.8). This conformation is called an *envelope form* because the structure resembles an opened envelope with the back flap raised. In the ribose moiety of most biomolecules, either C-2 or C-3 is out of the plane on the same side as C-5. These conformations are called C-2-endo and C-3-endo, respectively.



Chair form **Boat form**

FIGURE 11.7 Chair and boat forms of β-D-glucose. The chair form is more stable because hydrogen atoms occupy the axial positions, resulting in less steric hindrance. Abbreviations: a, axial; e, equatorial.

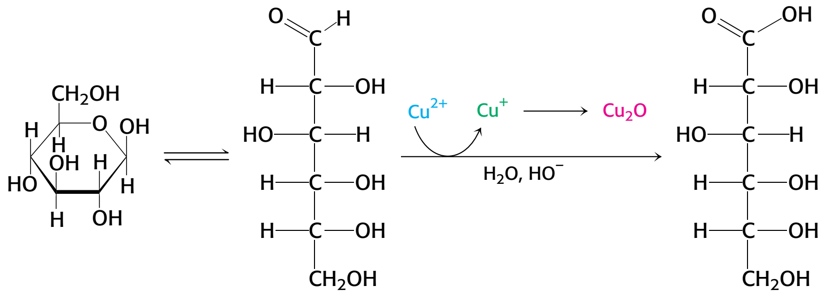


C-3-endo **C-2-endo**

FIGURE 11.8 Envelope conformations of β-D-ribose. The C-3-endo and C-2-endo forms of β-D-ribose are shown. The color indicates the four atoms that lie approximately in a plane.

Glucose is a reducing sugar

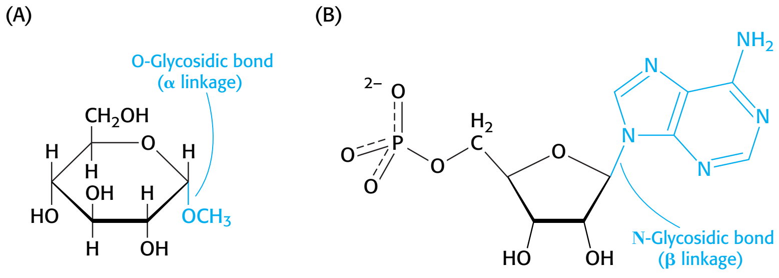
Because the α and β isomers of glucose are in an equilibrium that passes through the open-chain form, glucose has some of the chemical properties of free aldehydes, such as the ability to react with oxidizing agents. For example, glucose can react with cupric ion (Cu^{2+}), reducing it to cuprous ion (Cu^{+}), while being oxidized to gluconic acid.



Solutions of cupric ion (known as Fehling's solution) provide a simple test for the presence of sugars such as glucose. Sugars that react are called *reducing sugars;* those that do not are called *nonreducing sugars.* Reducing sugars can often nonspecifically react with a free amino group to form a stable covalent bond. For instance, as a reducing sugar, glucose reacts with hemoglobin to form glycosylated hemoglobin (hemoglobin A1c). Monitoring changes in the amount of glycosylated hemoglobin is an especially useful means of assessing the effectiveness of treatments for diabetes mellitus, a condition characterized by high levels of blood glucose (Section 27.3). Because the glycosylated hemoglobin remains in circulation, the amount of the modified hemoglobin corresponds to the long-term regulation—over several months—of glucose levels. In nondiabetic individuals, less than 6% of the hemoglobin is glycosylated, whereas, in uncontrolled diabetics, almost 10% of the hemoglobin is glycosylated. Although the glycosylation of hemoglobin has no effect on oxygen binding and is thus benign, similar reducing reactions are often detrimental because the glycosylations alter the normal biochemical function of the modified proteins. These modifcations, known as *advanced glycation end products* (AGE), have been implicated in aging, arteriosclerosis, and diabetes, as well as other pathological conditions.

Monosaccharides are joined to alcohols and amines through glycosidic bonds

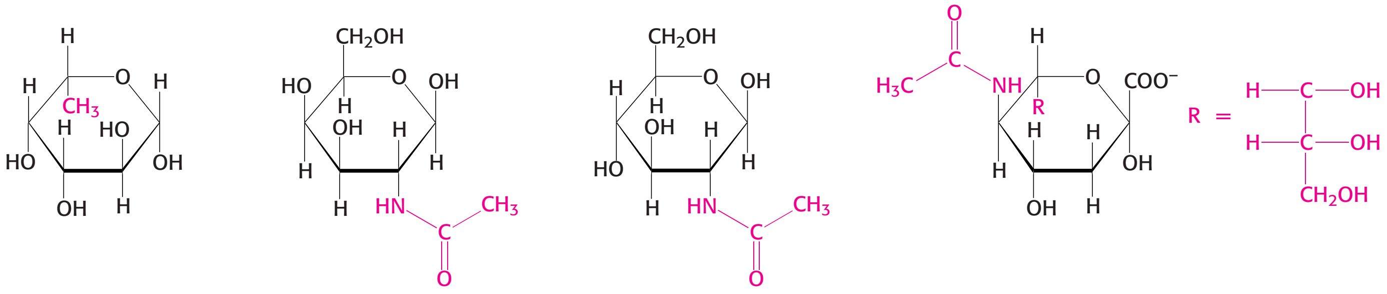
The biochemical properties of monosaccharides can by modified by reaction with other molecules. These modifications increase the biochemical versatility of carbohydrates, enabling them to serve as signal molecules or facilitating their metabolism. Three common reactants are alcohols, amines, and phosphates. A bond formed between the anomeric carbon atom of a carbohydrate and the oxygen atom of an alcohol is called a *glycosidic bond*—specifically, an O-*glycosidic bond*. *O*-Glycosidic bonds are prominent when carbohydrates are linked together to form long polymers and when they are attached to proteins (Figure 11.9). In addition, the anomeric carbon atom of a sugar can be linked to the nitrogen atom of an amine to form an N-*glycosidic bond*, such as when nitrogenous bases are attached to ribose units to form nucleosides. Carbohydrates can also be modifed by the attachment of functional groups to carbons other than the anomeric carbon (Figure 11.10).



α-D-Methylglucose

Adenosine monophosphate

FIGURE 11.9 O- and N-glycosidic linkages. (A) An O-glycosidic bond links glucose to a methyl group in α-D-methylglucose. (B) An N-glycosidic bond joins ribose to the base adenine in adenosine monophosphate.



β-L-Fucose (Fuc)

β-D-Acetylgalactosamine (GalNAc)

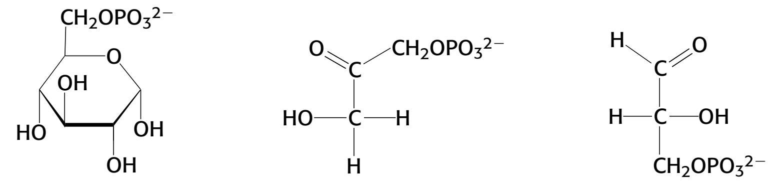
β-D-Acetylglucosamine (GlcNAc)

Sialic acid (Sia) (*N*-Acetylneuraminate)

FIGURE 11.10 Modified monosaccharides. Carbohydrates can be modified by the addition of substituents (shown in red) other than hydroxyl groups. Such modified carbohydrates are often expressed on cell surfaces.

Phosphorylated sugars are key intermediates in energy generation and biosyntheses

One sugar modification deserves special note because of its prominence in metabolism. The addition of phosphoryl groups is a common modification of sugars. For instance, the first step in the breakdown of glucose to obtain energy is its conversion into glucose 6-phosphate. Several subsequent intermediates in this metabolic pathway, such as dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, are phosphorylated sugars.



Glucose 6-phosphate (G-6P)

Dihydroxyacetone phosphate (DHAP)

Glyceraldehyde 3-phosphate (GAP)

Phosphorylation makes sugars anionic; the negative charge not only prevents these sugars from spontaneously leaving the cell by crossing lipid-bilayer membranes, but also prevents them from interacting with transporters of the unmodified sugar. Moreover, phosphorylation *creates reactive intermediates* that will more readily undergo metabolism. For example, a multiply phosphorylated derivative of ribose plays key roles in the biosyntheses of purine and pyrimidine nucleotides (Chapter 25).

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates

Because sugars contain hydroxyl groups, glycosidic bonds can join one monosaccharide to another. *Oligosaccharides* are built by the linkage of two or more monosaccharides by *O*-glycosidic bonds (Figure 11.11). In the disaccharide maltose, for example, two D-glucose residues are joined by a glycosidic linkage between the α-anomeric form of C-1 on one sugar and the hydroxyl oxygen atom on C-4 of the adjacent sugar. Such a linkage is called an α-1,4-glycosidic bond. Just as proteins have a directionality defined by the amino and carboxyl termini, oligosaccharides have a directionality defined by their reducing and nonreducing ends. The carbohydrate unit at the reducing end has a free anomeric carbon atom that has reducing activity because it can form the open-chain form, as discussed earlier (p. 321). By convention, this end of the oligosaccharide is still called the reducing end even when it is bound to another molecule such as a protein and thus no longer has reducing properties.

The fact that monosaccharides have multiple hydroxyl groups means that many different glycosidic linkages are possible. For example, consider three monosaccharides: glucose, mannose, and galactose. These molecules can be linked together in the laboratory to form more than 12,000 structures differing in the order of the monosaccharides and the hydroxyl groups participating in the glycosidic linkages. In this section, we will look at some of the most common oligosaccharides found in nature.

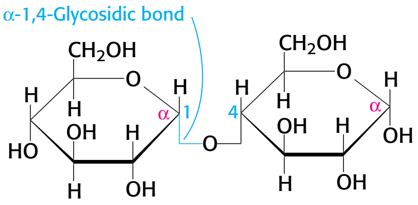
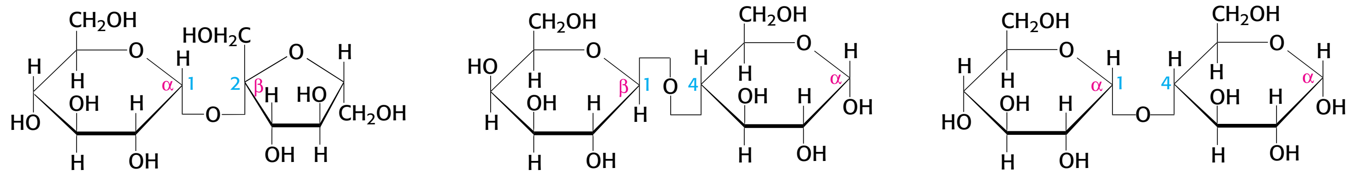


FIGURE 11.11 Maltose, a disaccharide. Two molecules of glucose are linked by an α-1,4-glycosidic bond to form the disaccharide maltose. The angles in the bonds to the central oxygen atom do not denote carbon atoms. The angles are added only for ease of illustration. The glucose molecule on the right is capable of assuming the open-chain form, which can act as a reducing agent. The glucose molecule on the left cannot assume the open-chain form, because the C-1 carbon atom is bound to another molecule.

Sucrose, lactose, and maltose are the common disaccharides

A *disaccharide* consists of two sugars joined by an *O*-glycosidic bond. Three abundant disaccharides that we encounter frequently are sucrose, lactose, and maltose (Figure 11.12). *Sucrose* (common table sugar) is obtained commercially from sugar cane or sugar beets. The anomeric carbon atoms of a glucose unit and a fructose unit are joined in this disaccharide; the configuration of this glycosidic linkage is α for glucose and β for fructose. Sucrose can be cleaved into its component monosaccharides by the enzyme *sucrase*. *Lactose*, the disaccharide of milk, consists of galactose joined to



Sucrose
(α-D-Glucopyranosyl-(1 → 2)-β-D-fructofuranose

Lactose
(β-D-Galactopyranosyl-(1 → 4)-α-D-glucopyranose

Maltose
(α-D-Glucopyranosyl-(1 → 4)-α-D-glucopyranose

FIGURE 11.12 **Common disaccharides.** Sucrose, lactose, and maltose are common dietary components. As in Figure 11.11, the angles in the bonds to the central oxygen atoms do not denote carbon atoms.

glucose by a β-1,4-glycosidic linkage. Lactose is hydrolyzed to these monosaccharides by *lactase* in human beings and by *β-galactosidase* in bacteria. In *maltose,* two glucose units are joined by an α-1,4-glycosidic linkage. Maltose comes from the hydrolysis of large polymeric oligosaccharides such as starch and glycogen and is in turn hydrolyzed to glucose by *maltase*. Sucrase, lactase, and maltase are located on the outer surfaces of epithelial cells lining the small intestine. The cleavage products of sucrose, lactose, and maltose can be further processed to provide energy in the form of ATP.

Glycogen and starch are storage forms of glucose

Glucose is an important energy source in virtually all life forms. However, free glucose molecules cannot be stored because in high concentrations, glucose will disturb the osmotic balance of the cell, potentially resulting in cell death. The solution is to store glucose as units in a large polymer, which is not osmotically active.

Large polymeric oligosaccharides, formed by the linkage of multiple monosaccharides, are called *polysaccharides* and play vital roles in energy storage and in maintaining the structural integrity of an organism. If all of the monosaccharide units in a polysaccharide are the same, the polymer is called a *homopolymer*. The most common homopolymer in animal cells is *glycogen,* the storage form of glucose. Glycogen is present in most of our tissues but is most abundant in muscle and liver. As will be considered in detail in Chapter 21, glycogen is a large, branched polymer of glucose residues. Most of the glucose units in glycogen are linked by α-1,4-glycosidic bonds. The branches are formed by α-1,6-glycosidic bonds, present about once in 10 units (Figure 11.13).

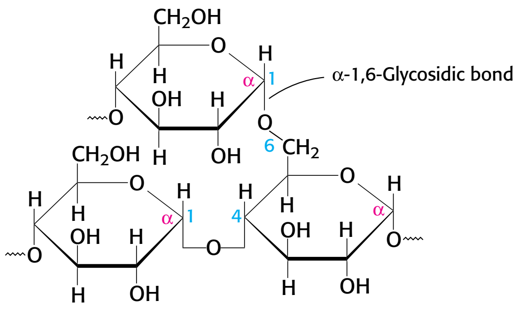
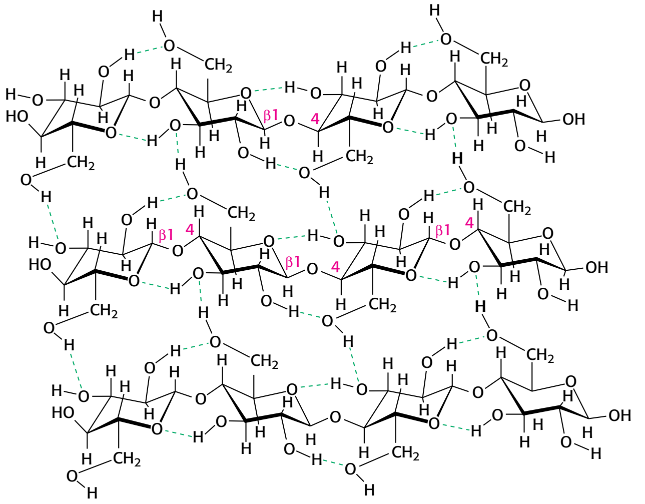


FIGURE 11.13 **Branch point in glycogen.** Two chains of glucose molecules joined by α-1,4-glycosidic bonds are linked by an α-1,6-glycosidic bond to create a branch point. Such an α-1,6-glycosidic bond forms at approximately every 10 glucose units, making glycogen a highly branched molecule.

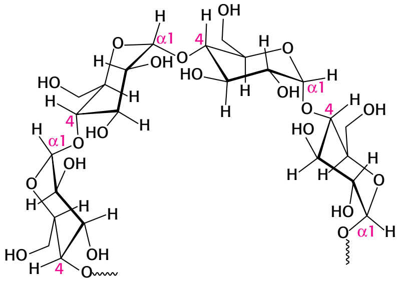
The nutritional reservoir in plants is the homopolymer *starch,* of which there are two forms. *Amylose,* the unbranched type of starch, consists of glucose residues in α-1,4 linkage. *Amylopectin,* the branched form, has about 1 α-1,6 linkage per 30 α-1,4 linkages, in similar fashion to glycogen except for its lower degree of branching. More than half the carbohydrate ingested by human beings is starch found in wheat, potatoes, and rice, to name just a few sources. Amylopectin, amylose, and glycogen are rapidly hydrolyzed by *α-amylase,* an enzyme secreted by the salivary glands and the pancreas.

Cellulose, a structural component of plants, is made of chains of glucose

Cellulose, the other major polysaccharide of glucose found in plants, serves a structural rather than a nutritional role as an important component of the plant cell wall. *Cellulose is among the most abundant organic compounds in the biosphere.* Some 10^{15} kg of cellulose is synthesized and degraded on Earth each year, an amount 1000 times as great as the



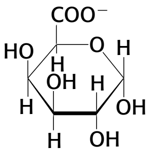
Cellulose
(β-1,4 linkages)



Starch and glycogen
(α-1,4 linkages)

FIGURE 11.14 Glycosidic bonds determine polysaccharide structure. The β-1,4 linkages favor straight chains, which are optimal for structural purposes. The α-1,4 linkages favor bent structures, which are more suitable for storage.

combined weight of the human race. Cellulose is an unbranched polymer of glucose residues joined by β-1,4 linkages, in contrast with the α-1,4 linkage seen in starch and glycogen. This simple difference in stereochemistry yields two molecules with very different properties and biological functions. The β configuration allows cellulose to form very long, straight chains. Fibrils are formed by parallel chains that interact with one another through hydrogen bonds, generating a rigid, supportive structure. The straight chains formed by β linkages are optimal for the construction of fibers having a high tensile strength. The α-1,4 linkages in glycogen and starch produce a very different molecular architecture: a hollow helix is formed instead of a straight chain (Figure 11.14). The hollow helix formed by α linkages is well suited to the formation of a more-compact, accessible store of sugar. Although mammals lack cellulases and therefore cannot digest wood and vegetable fibers, cellulose and other plant fibers are still an important constituent of the mammalian diet as a component of dietary fiber. Soluble fiber such as *pectin* (polygalacturonic acid) slows the movement of food through the gastrointestinal tract, allowing improved digestion and the absorption of nutrients. Insoluble fibers, such as cellulose, increase the rate at which digestion products pass through the large intestine. This increase in rate can minimize exposure to toxins in the diet.

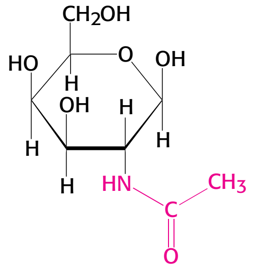


Galacturonic acid

We have considered only homopolymers of glucose. However, given the variety of different monosaccharides that can be put together in any number of arrangements, the number of possible polysaccharides is huge. We will consider some of these polysaccharides shortly.

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins

A carbohydrate group can be covalently attached to a protein to form a *glycoprotein.* Such modifications are not rare, as 50% of the proteome consists of glycoproteins. We will examine three classes of glycoproteins. The first class is simply referred to as glycoproteins. In glycoproteins of this class, the protein constituent is the largest component by weight. This versatile class plays a variety of biochemical roles. Many glycoproteins



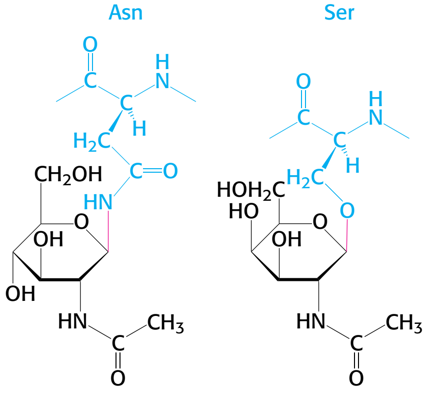
β-D-Acetylgalactosamine (GalNAc)

are components of cell membranes, where they take part in processes such as cell adhesion and the binding of sperm to eggs. Other glycoproteins are formed by linking carbohydrates to soluble proteins. Many of the proteins secreted from cells are glycosylated, or modified by the attachment of carbohydrates, including most proteins present in the serum component of blood.

The second class of glycoproteins comprises the *proteoglycans*. The protein component of proteoglycans is conjugated to a particular type of polysaccharide called a *glycosaminoglycan*. Carbohydrates make up a much larger percentage by weight of the proteoglycan compared with simple glycoproteins. Proteoglycans function as structural components and lubricants.

Mucins, or *mucoproteins,* are, like proteoglycans, predominantly carbohydrate. *N*-Acetylgalactosamine is usually the carbohydrate moiety bound to the protein in mucins. *N*-Acetylgalactosamine is an example of an *amino sugar,* so named because an amino group replaces a hydroxyl group. Mucins, a key component of mucus, serve as lubricants.

Glycosylation greatly increases the complexity of the proteome. A given protein with several potential glycosylation sites can have many different glycosylated forms (called *glycoforms*), each of which can be generated only in a specific cell type or developmental stage.



***N*-linked GlcNAc** ***O*-linked GalNAc**

FIGURE 11.15 Glycosidic bonds between proteins and carbohydrates. A glycosidic bond links a carbohydrate to the side chain of asparagine (*N*-linked) or to the side chain of serine or threonine (*O*-linked). The glycosidic bonds are shown in red.

Carbohydrates can be linked to proteins through asparagine (*N*-linked) or through serine or threonine (*O*-linked) residues

Sugars in glycoproteins are attached either to the amide nitrogen atom in the side chain of asparagine (termed an N-*linkage*) or to the oxygen atom in the side chain of serine or threonine (termed an O-*linkage*), as shown in Figure 11.15. An asparagine residue can accept an oligosaccharide only if the residue is part of an Asn-X-Ser or Asn-X-Thr sequence, in which X can be any residue, except proline. However, not all potential sites are glycosylated. Which sites are glycosylated depends on other aspects of the protein structure and on the cell type in which the protein is expressed. All *N*-linked oligosaccharides have in common a pentasaccharide core consisting of three mannose and two *N*-acetylglucosamine residues. Additional sugars are attached to this core to form the great variety of oligosaccharide patterns found in glycoproteins (Figure 11.16).

Abbreviations for sugars

Fuc	▲	Fucose
Gal	●	Galactose
GalNAc	■	*N*-Acetylgalactosamine
Glc	●	Glucose
GlcNAc	■	*N*-Acetylglucosamine
Man	●	Mannose
Sia	◆	Sialic acid

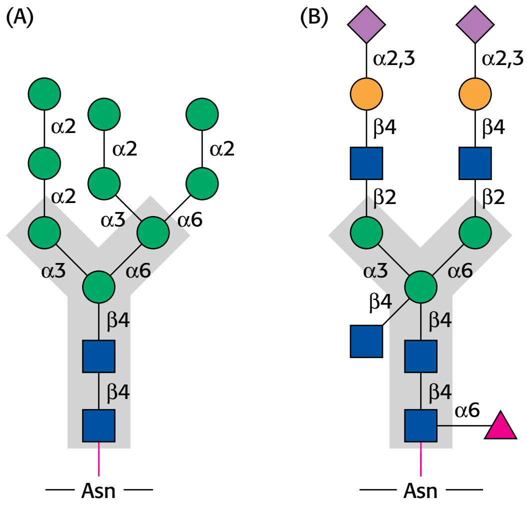


FIGURE 11.16 *N*-linked oligosaccharides. A pentasaccharide core (shaded gray) is common to all *N*-linked oligosaccharides and serves as the foundation for a wide variety of *N*-linked oligosaccharides, two of which are illustrated: (A) high-mannose type; (B) complex type.

The glycoprotein erythropoietin is a vital hormone

Let us look at a glycoprotein present in the blood serum that has dramatically improved treatment for anemia, particularly that induced by cancer chemotherapy. The glycoprotein hormone *erythropoietin* (EPO) is secreted by the kidneys and stimulates the production of red blood cells. EPO is composed of 165 amino acids and is *N*-glycosylated at three asparagine residues and *O*-glycosylated on a serine residue (Figure 11.17). The mature EPO is 40% carbohydrate by weight, and glycosylation enhances the stability of the protein in the blood. Unglycosylated protein has only about 10% of the bioactivity of the glycosylated form because the protein is rapidly removed from the blood by the kidneys. The availability of recombinant human EPO has greatly aided the treatment of anemias. However, some endurance athletes have used recombinant human EPO to increase the red-blood-cell count and hence their oxygen-carrying capacity. Drug-testing laboratories are able to distinguish some forms of prohibited human recombinant EPO from natural EPO in athletes by detecting differences in their glycosylation patterns through the use of isoelectric focusing (p. 73).

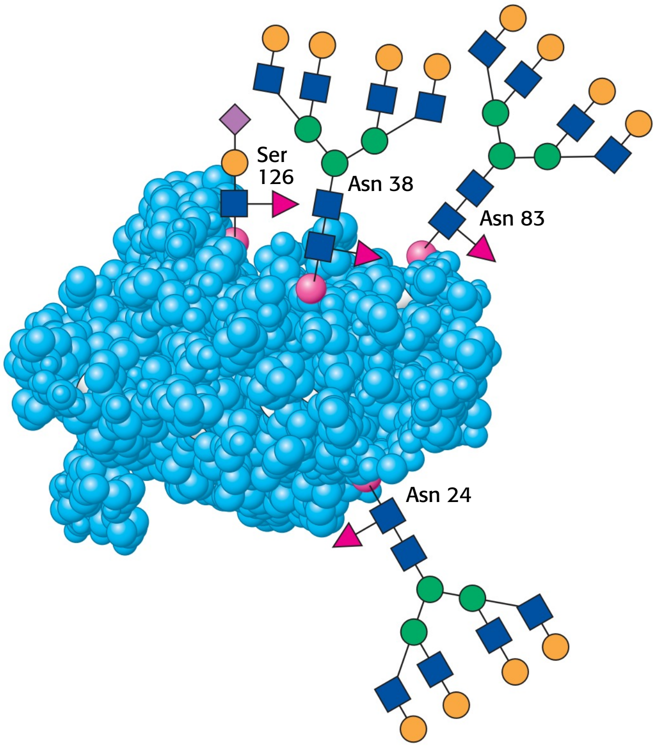
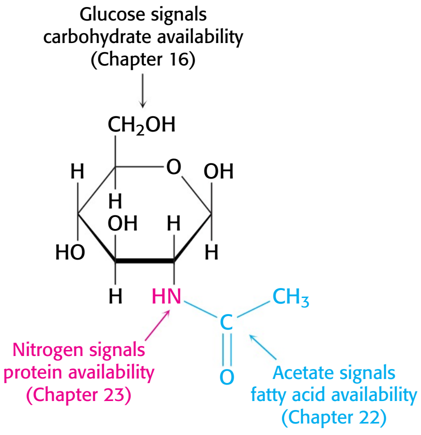


FIGURE 11.17 Oligosaccharides attached to erythropoietin. Erythropoietin has oligosaccharides linked to three asparagine residues and one serine residue. The structures shown are approximately to scale. See Figure 11.16 for the carbohydrate key. [Drawn from 1BUY.pdf.]

Glycosylation functions in nutrient sensing

An especially important glycosylation reaction is the covalent attachment of N-acetylglucosamine (GlcNAc) to serine or threonine residues of cellular proteins, a reaction catalyzed by *GlcNAc transferase*. The concentration of GlcNAc reflects the active metabolism of carbohydrates, amino acids and fats, indicating that nutrients are abundant (Figure 11.18). More than one thousand proteins are modified by GlcNAcylation, including transcription factors and components of signaling pathways. Interestingly, because the GlcNAcylation sites are also potential phosphorylation sites, O-GlcNAc transferase and protein kinases may be involved in cross talk to modulate one another's signaling activity. Like phosphorylation, GlcNAcylation is reversible, with *GlcNAcase* catalyzing the removal of the carbohydrate. Dysregulation of GlcNAc transferase has been linked to insulin resistance, diabetes, cancer and neurological pathologies.



β-D-Acetylglucosamine
(GlcNAc)

FIGURE 11.18 Glycosylation as a nutrient sensor. N-acetylglucosamine is attached to proteins when nutrients are abundant.

Proteoglycans, composed of polysaccharides and protein, have important structural roles

As stated earlier, proteoglycans are proteins attached to glycosaminoglycans. The glycosaminoglycan makes up as much as 95% of the biomolecule by weight, and so the proteoglycan resembles a polysaccharide more than a protein. Proteoglycans not only function as lubricants and structural components in connective tissue, but also mediate the adhesion of cells to the extracellular matrix, and bind factors that regulate cell proliferation.

The properties of proteoglycans are determined primarily by the glycosaminoglycan component. Many glycosaminoglycans are made of repeating units of disaccharides containing a derivative of an

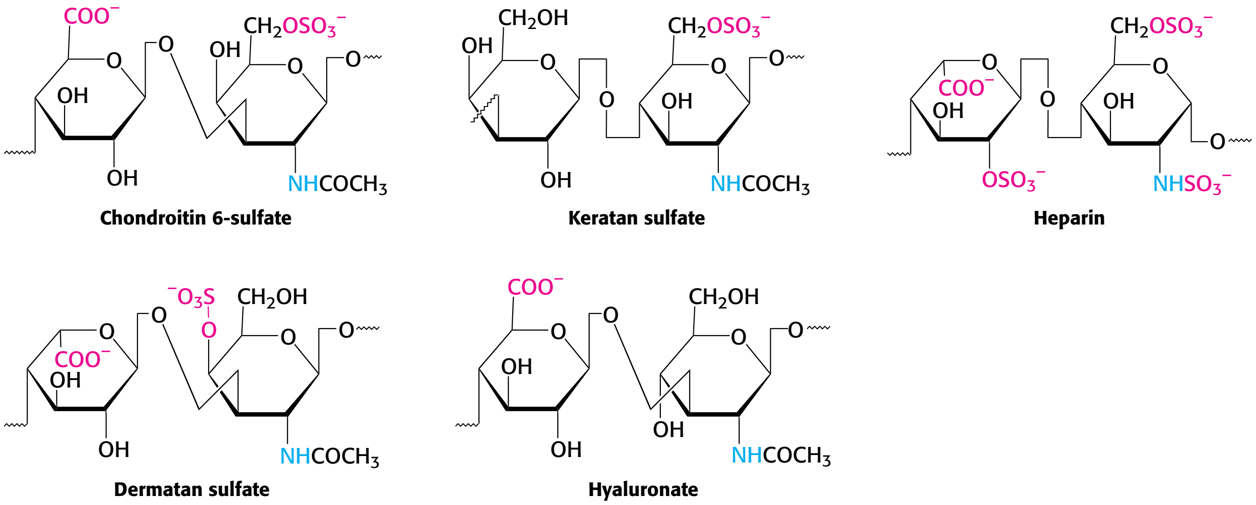


FIGURE 11.19 Repeating units in glycosaminoglycans. Structural formulas for five repeating units of important glycosaminoglycans illustrate the variety of modifications and linkages that is possible. Amino groups are shown in blue and negatively charged groups in red. Hydrogen atoms have been omitted for clarity. The right-hand structure is a glucosamine derivative in each case.

amino sugar, either glucosamine or galactosamine (Figure 11.19). At least one of the two sugars in the repeating unit has a *negatively charged carboxylate or sulfate group*. The major glycosaminoglycans in animals are chondroitin sulfate, keratan sulfate, heparin, dermatan sulfate, and hyaluronate. Recall that heparin acts as an anticoagulant to assist the termination of blood clotting (p. 307). *Mucopolysaccharidoses* are a collection of diseases, such as Hurler disease, that result from the inability to degrade glycosaminoglycans (Figure 11.20). Although precise clinical features vary with the disease, all mucopolysaccharidoses result in skeletal deformities and reduced life expectancies.

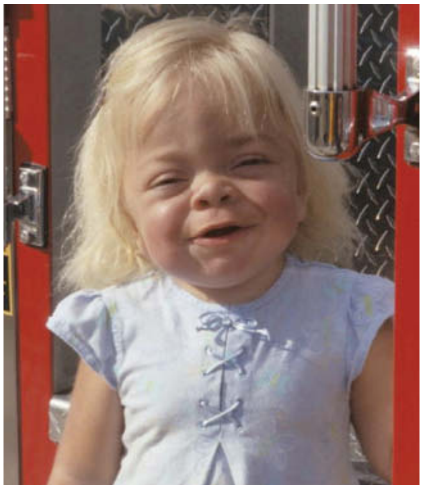


FIGURE 11.20 Hurler disease. Formerly called gargoylism, Hurler disease is a mucopolysaccharidosis having symptoms that include wide nostrils, a depressed nasal bridge, thick lips and earlobes, and irregular teeth. In Hurler disease, glycosaminoglycans cannot be degraded. The excess of these molecules are stored in the soft tissue of the facial regions, resulting in the characteristic facial features. [Courtesy National MPS Society, www.mpssociety.org.]

Proteoglycans are important components of cartilage

Among the best-characterized members of this diverse class is the proteoglycan in the extracellular matrix of cartilage. The proteoglycan *aggrecan* and the protein *collagen* are key components of cartilage. The triple helix of collagen (p. 45) provides structure and tensile strength, whereas aggrecan serves as a shock absorber. The protein component of aggrecan is a large molecule composed of 2397 amino acids. The protein has three globular domains, and the site of glycosaminoglycan attachment is the extended region between globular domains 2 and 3. This linear region contains highly repetitive amino acid sequences, which are sites for the attachment of keratan sulfate and chondroitin sulfate. Many molecules of aggrecan are in turn noncovalently bound through the first globular domain to a very long filament formed by linking together molecules of the glycosaminoglycan hyaluronate (Figure 11.21). Water is bound to the glycosaminoglycans, attracted by the many negative charges. Aggrecan can cushion compressive forces because the absorbed water enables it to spring back after having been deformed. When pressure is exerted, as when the foot hits the ground while walking, water is squeezed from the glycosaminoglycan, cushioning the impact. When the pressure is released, the water rebinds. *Osteoarthritis*, the most common form of arthritis, results when water is lost from proteoglycan with aging. Other forms of arthritis can result from the proteolytic degradation of aggrecan and collagen in the cartilage.

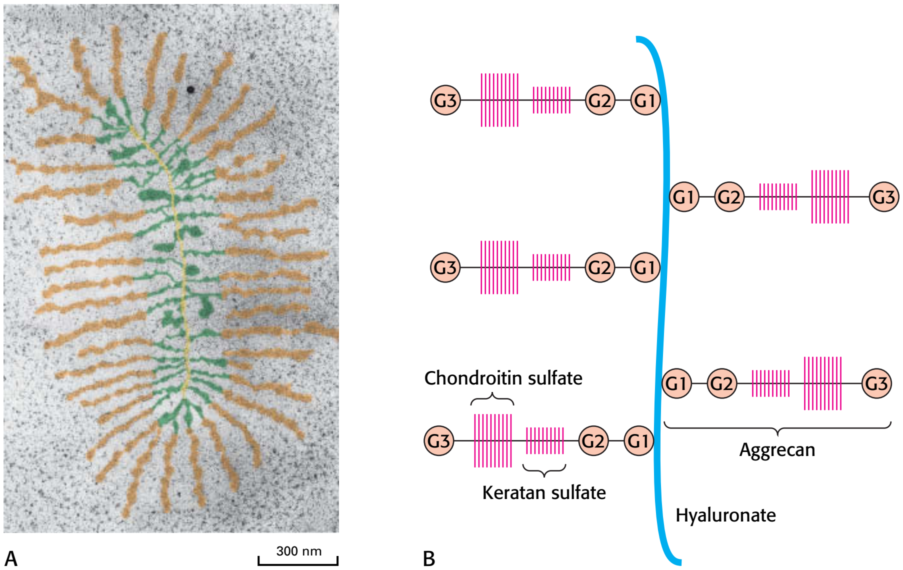


FIGURE 11.21 Structure of proteoglycan from cartilage. (A) Electron micrograph of a proteoglycan from cartilage (with false color added). Proteoglycan monomers emerge laterally at regular intervals from opposite sides of a central filament of hyaluronate. (B) Schematic representation. G = globular domain. [(A) Courtesy of Dr. Lawrence Rosenberg. From J. A. Buckwalter and L. Rosenberg. *Collagen Relat*. *Res*. 3:489–504, 1983.]

FIGURE 11.22 Chitin, a glycosaminoglycan, is present in insect wings and the exoskeleton. Glycosaminoglycans are components of the exoskeletons of insects, crustaceans, and arachnids. [FLPA/Alamy.]

In addition to being a key component of structural tissues, glycosaminoglycans are common throughout the biosphere. Chitin is a glycosaminoglycan found in the exoskeleton of insects, crustaceans, and arachnids and is, next to cellulose, the second most abundant polysaccharide in nature (Figure 11.22). Cephalopods such as squid use their razor sharp beaks, which are made of extensively crosslinked chitin, to disable and consume prey.

Mucins are glycoprotein components of mucus

A third class of glycoproteins is the *mucins* (mucoproteins). In mucins, the protein component is extensively glycosylated at serine or threonine residues by N-acetylgalactosamine (Figure 11.10). Mucins are capable of forming large polymeric structures and are common in mucous secretions. These glycoproteins are synthesized by specialized cells in the tracheobronchial, gastrointestinal, and genitourinary tracts. Mucins are abundant in saliva where they function as lubricants.

A model of a mucin is shown in Figure 11.23A. The defining feature of the mucins is a region of the protein backbone termed the *variable number of tandem repeats* (VNTR) region, which is rich in serine and threonine residues that are O-glycosylated. Indeed, the carbohydrate moiety can account for as much as 80% of the molecule by weight. A number of core carbohydrate structures are conjugated to the protein component of mucin. Figure 11.23B shows one such structure.

Mucins adhere to epithelial cells and act as a protective barrier; they also hydrate the underlying cells. In addition to protecting cells from environmental insults, such as stomach acid, inhaled chemicals in the lungs, and bacterial infections, mucins have roles in fertilization, the immune response, and cell adhesion. Mucins are overexpressed in bronchitis and cystic fibrosis, and the overexpression of mucins is characteristic of adenocarcinomas—cancers of the glandular cells of epithelial origin.

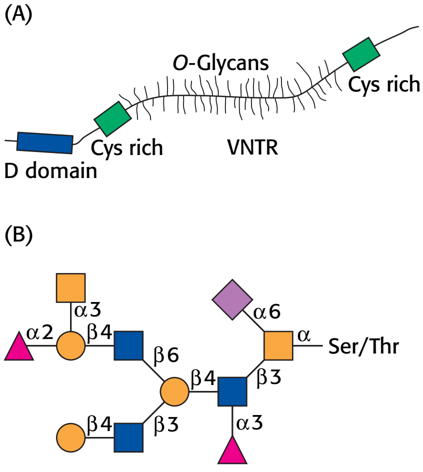


FIGURE 11.23 Mucin structure. (A) A schematic representation of a mucoprotein. The VNTR region is highly glycosylated, forcing the molecule into an extended conformation. The Cys-rich domains and the D domain facilitate the polymerization of many such molecules. (B) An example of an oligosaccharide that is bound to the VNTR region of the protein. See Figure 11.16 for the carbohydrate key. [Information from A. Varki et al. (Eds.), *Essentials of Glycobiology*, 2d ed. (Cold Spring Harbor Press, 2009), pp. 117, 118.]

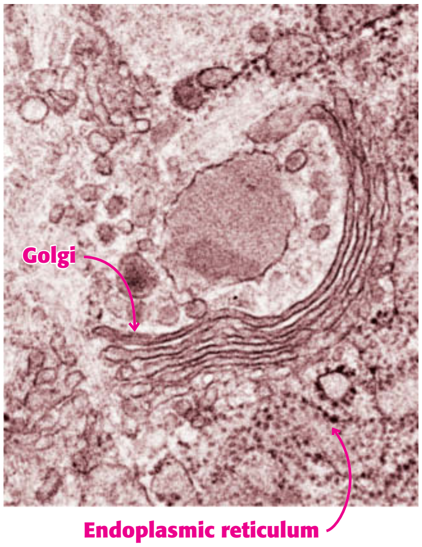
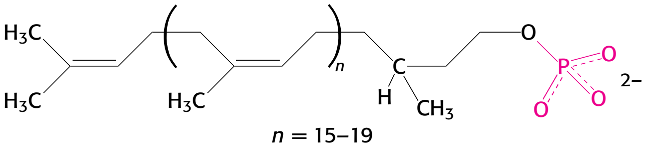


FIGURE 11.24 Golgi complex and endoplasmic reticulum. The electron micrograph shows the Golgi complex and adjacent endoplasmic reticulum. The black dots on the cytoplasmic surface of the ER membrane are ribosomes. [Micrograph courtesy of Lynne Mercer.]

Protein glycosylation takes place in the lumen of the endoplasmic reticulum and in the Golgi complex

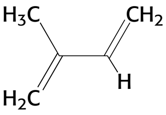
The major pathway for protein glycosylation takes place inside the lumen of the *endoplasmic reticulum* (ER) and in the *Golgi complex,* organelles that play central roles in protein trafficking (Figure 11.24). The protein is synthesized by ribosomes attached to the cytoplasmic face of the ER membrane, and the peptide chain is inserted into the lumen of the ER (Section 30.6). The *N*-linked glycosylation begins in the ER and continues in the Golgi complex, whereas the *O*-linked glycosylation takes place exclusively in the Golgi complex.

A large oligosaccharide destined for attachment to the asparagine residue of a protein is assembled on *dolichol phosphate,* a specialized lipid molecule located in the ER membrane and containing about 20 isoprene (C_5) units.



n = 15–19

Dolichol phosphate



Isoprene

The terminal phosphate group of the dolichol phosphate is the site of attachment of the oligosaccharide. This activated (energy-rich) form of the oligosaccharide is subsequently transferred to a specific asparagine residue of the growing polypeptide chain by an enzyme located on the lumenal side of the ER.

Proteins in the lumen of the ER and in the ER membrane are transported to the Golgi complex, which is a stack of flattened membranous sacs. *Carbohydrate units of glycoproteins are altered and elaborated in the Golgi complex.* The *O*-linked sugar units are fashioned there, and the *N*-linked sugars, arriving from the ER as a component of a glycoprotein, are modified in many different ways. *The Golgi complex is the major sorting center of the cell.* Proteins proceed from the Golgi complex to lysosomes, secretory granules, or the plasma membrane, according to signals encoded within their amino acid sequences and three-dimensional structures (Figure 11.25).

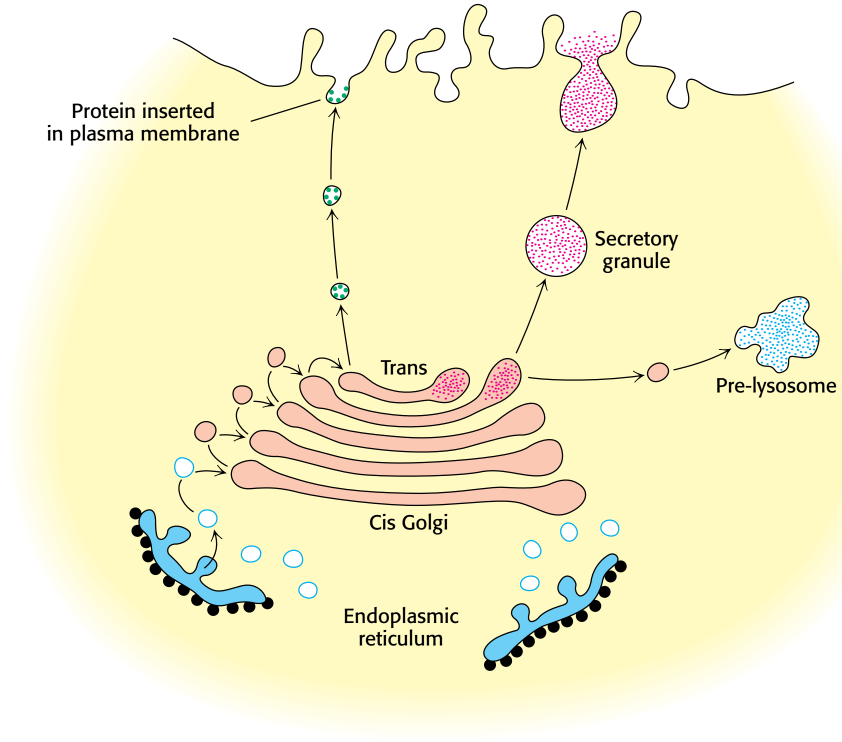


FIGURE 11.25 Golgi complex as sorting center. The Golgi complex is the sorting center in the targeting of proteins to lysosomes, secretory vesicles, and the plasma membrane. The cis face of the Golgi complex receives vesicles from the endoplasmic reticulum, and the trans face sends a different set of vesicles to target sites. Vesicles also transfer proteins from one compartment of the Golgi complex to another. [Courtesy of Dr. Marilyn Farquhar.]

Specific enzymes are responsible for oligosaccharide assembly

How are the complex carbohydrates formed, be they unconjugated molecules such as glycogen or components of glycoproteins? Complex carbohydrates are synthesized through the action of specific enzymes, *glycosyltransferases*, which catalyze the formation of glycosidic bonds. Given the diversity of known glycosidic linkages, many different enzymes are required. Indeed, glycosyltransferases account for 1% to 2% of gene products in all organisms examined.

While dolichol phosphate-linked oligosaccharides are substrates for some glycosyltransferases, the most common carbohydrate donors for glycosyltransferases are activated sugar nucleotides, such as UDP-glucose (UDP is the abbreviation for uridine diphosphate) (Figure 11.26). The attachment of a nucleotide to enhance the energy content of a molecule is a common strategy in biosynthesis that we will see many times in our study of biochemistry. The acceptor substrates for glycosyltransferases are quite varied and include carbohydrates, serine, threonine, and asparagine residues of proteins, lipids, and even nucleic acids.

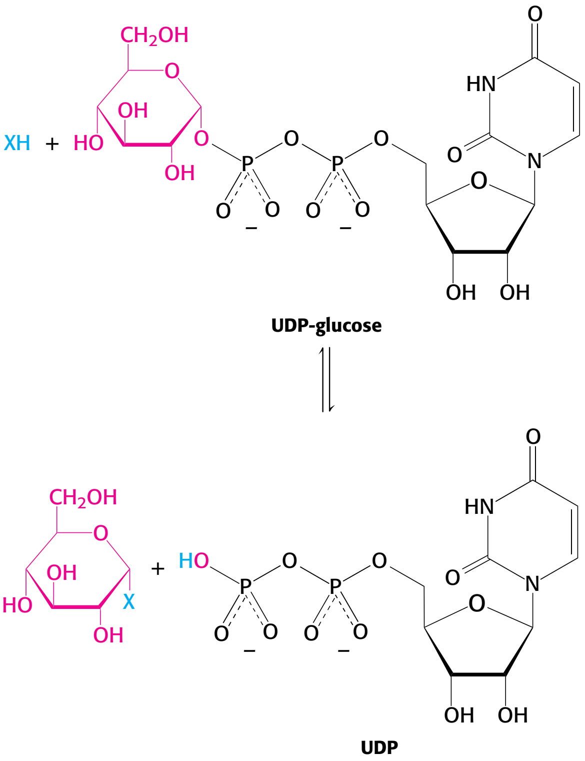


FIGURE 11.26 **General form of a glycosyltransferase reaction.** The sugar to be added comes from a sugar nucleotide—in this case, UDP-glucose. The acceptor, designated X in this illustration, can be one of a variety of biomolecules, including other carbohydrates or proteins.

Blood groups are based on protein glycosylation patterns

The human ABO blood groups illustrate the effects of glycosyltransferases on the formation of glycoproteins. Each blood group is designated by the presence of one of the three different carbohydrates, termed A, B, or O, attached to glycoproteins and glycolipids on the surfaces of red blood cells (Figure 11.27). These structures have in common an oligosaccharide foundation called the O (or sometimes H) antigen. The A and B antigens differ from the O antigen by the addition of one extra monosaccharide, either *N*-acetylgalactosamine (for A) or galactose (for B) through an α-1,3 linkage to a galactose moiety of the O antigen.

Specific glycosyltransferases add the extra monosaccharide to the O antigen. Each person inherits the gene for one glycosyltransferase of this type from each parent. The type A transferase specifically adds *N*-acetylgalactosamine, whereas the type B transferase adds galactose. These enzymes are identical in all but 4 of 354 positions. The O phenotype is the result of a mutation in the O transferase that results in the synthesis of an inactive enzyme.

These structures have important implications for blood transfusions and other transplantation procedures. If an antigen not normally present in a person is introduced, the person's immune system recognizes it as foreign. Red-blood-cell lysis occurs rapidly, leading to a severe drop in blood pressure (hypotension), shock, kidney failure, and death from circulatory collapse.

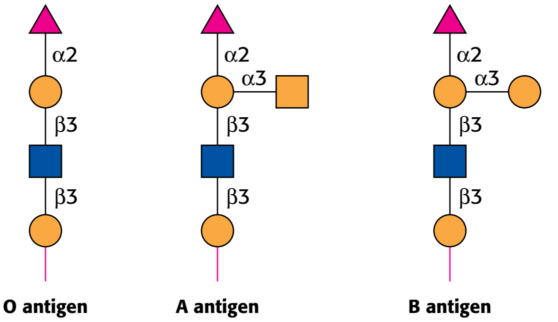


FIGURE 11.27 **Structures of A, B, and O oligosaccharide antigens.** The carbohydrate structures shown are depicted symbolically by employing a scheme (see the key in Figure 11.16) that is becoming widely used.

Why are different blood types present in the human population? Suppose that a pathogenic organism such as a parasite expresses on its cell surface a carbohydrate antigen similar to one of the blood-group antigens. This antigen may not be readily detected as foreign in a person whose blood type matches the parasite antigen, and the parasite will flourish. However, other people with different blood types will be protected. Hence, there will be selective pressure on human beings to vary blood type to prevent parasitic mimicry and a corresponding selective pressure on parasites to enhance mimicry. The constant "arms race" between pathogenic microorganisms and human beings drives the evolution of diversity of surface antigens within the human population.

Errors in glycosylation can result in pathological conditions

Although the role of carbohydrate attachment to proteins is not known in detail in most cases, data indicate that this glycosylation is important for the processing and stability of these proteins, as it is for EPO. Certain types of muscular dystrophy can be traced to improper glycosylation of dystroglycan, a membrane protein that links the extracellular matrix with the cytoskeleton. Indeed, an entire family of severe inherited human diseases called *congenital disorders of glycosylation* has been identified. These pathological conditions reveal the importance of proper modification of proteins by carbohydrates and their derivatives.

An especially clear example of the role of glycosylation is provided by *I-cell disease* (also called *mucolipidosis II*), a lysosomal storage disease. Normally, a carbohydrate marker directs certain digestive enzymes from the Golgi complex to lysosomes. *Lysosomes* are organelles that degrade and recycle damaged cellular components or material brought into the cell by endocytosis. In patients with I-cell disease, lysosomes contain large *inclusions* of undigested glycosaminoglycans and glycolipids—hence the "I" in the name of the disease. These inclusions are present because the enzymes normally responsible for the degradation of glycosaminoglycans are missing from affected lysosomes. Remarkably, the enzymes are present at very high levels in the blood and urine. Thus, active enzymes are synthesized, but, in the absence of appropriate glycosylation, they are exported instead of being sequestered in lysosomes. In other words, *in I-cell disease, a whole series of enzymes are incorrectly addressed and delivered to the wrong location.* Normally, these enzymes contain a mannose 6-phosphate residue as a component of an *N*-oligosaccharide that serves as the marker directing the enzymes from the Golgi complex to lysosomes. In I-cell disease, however, the attached mannose lacks a phosphate. I-cell patients are deficient in the *N-acetylglucosamine phosphotransferase* catalyzing the first step in the addition of the phosphoryl group; the consequence is the mistargeting of eight essential enzymes (Figure 11.28). I-cell disease causes the patient to suffer severe psychomotor retardation and skeletal deformities, similar to those in Hurler disease. Remarkably, mutations in the phosphotransferase have also been linked to stuttering. Why some mutations cause stuttering while other cause I-cell disease is a mystery.

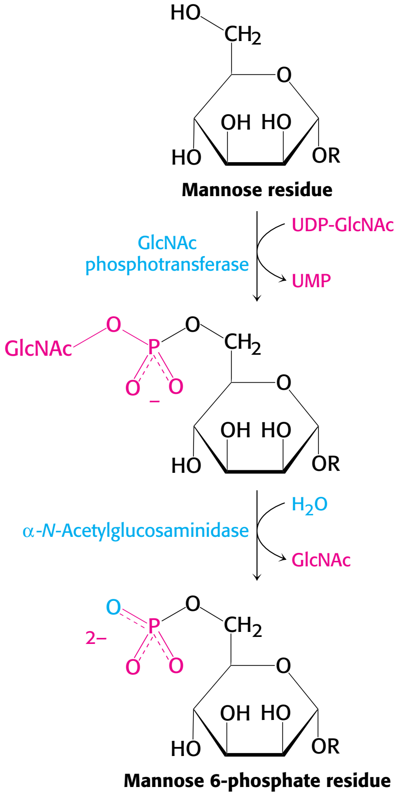

HO
CH2
O
OH HO
HO
OR
Mannose residue
GlcNAc phosphotransferase
UDP-GlcNAc
UMP
GlcNAc
O
P
O
O
O
–
CH2
α-N-Acetylglucosaminidase
H2O
GlcNAc
2–
Mannose 6-phosphate residue

FIGURE 11.28 Formation of a mannose 6-phosphate marker. A glycoprotein destined for delivery to lysosomes acquires a phosphate marker in the Golgi compartment in a two-step process. First, GlcNAc phosphotransferase adds a phospho-*N*-acetylglucosamine unit to the 6-OH group of a mannose, and then an *N*-acetylglucosaminidase removes the added sugar to generate a mannose 6-phosphate residue in the core oligosaccharide.

Oligosaccharides can be "sequenced"

How is it possible to determine the structure of a glycoprotein—the oligosaccharide structures and their points of attachment? Most approaches make use of enzymes that cleave oligosaccharides at specific types of linkages.

The first step is to detach the oligosaccharide from the protein. For example, *N*-linked oligosaccharides can be released from proteins by an enzyme such as *peptide N-glycosidase F,* which cleaves the *N*-glycosidic bonds linking the oligosaccharide to the protein. The oligosaccharides can then be isolated and analyzed. Matrix-assisted laser desorption/ionization/time-of-flight (MALDI-TOF) or other mass spectrometric techniques (Section 3.3) provide the mass of an oligosaccharide fragment. However, many possible oligosaccharide structures are consistent with a given mass. More-complete information can be obtained by cleaving the oligosaccharide with enzymes of varying specificities. For example, *β-1,4-galactosidase*

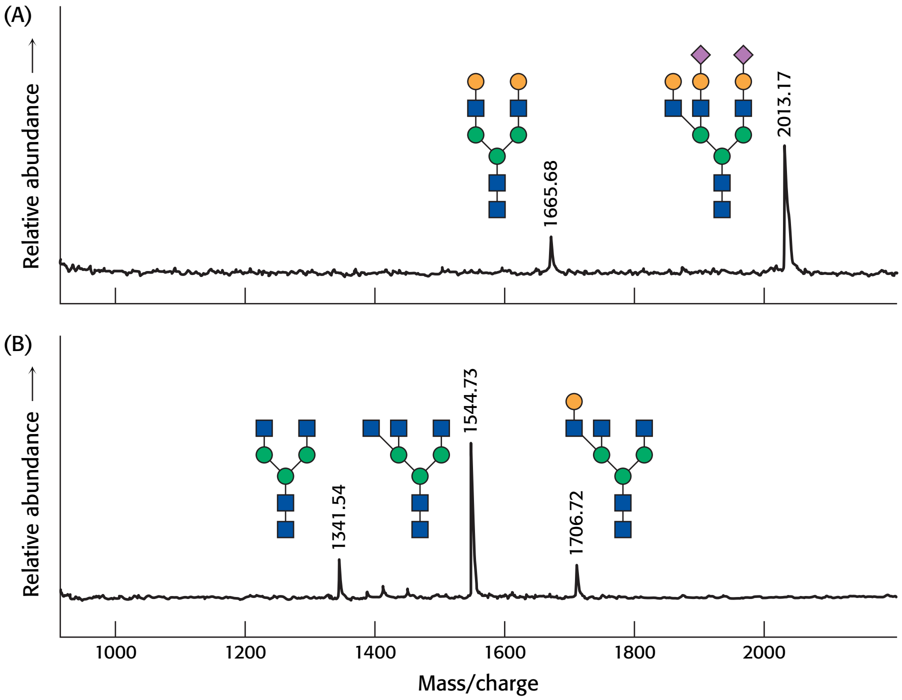


FIGURE 11.29 Mass spectrometric "sequencing" of oligosaccharides. Carbohydrate-cleaving enzymes were used to release and specifically cleave the oligosaccharide component of the glycoprotein fetuin from bovine serum. Parts A and B show the masses obtained with MALDI-TOF spectrometry as well as the corresponding structures of the oligosaccharide-digestion products: (A) digestion with peptide *N*-glycosidase F (to release the oligosaccharide from the protein) and neuraminidase; (B) digestion with peptide *N*-glycosidase F, neuraminidase, and β-1,4-galactosidase. Knowledge of the enzyme specificities and the masses of the products permits the characterization of the oligosaccharide. See Figure 11.16 for the carbohydrate key. [Data from A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, G. W. Hart, and J. Marth (Eds.), *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 1999), p. 596.]

cleaves β-glycosidic bonds exclusively at galactose residues. The products can again be analyzed by mass spectrometry (Figure 11.29). The repetition of this process with the use of an array of enzymes of different specificity will eventually reveal the structure of the oligosaccharide.

Proteases applied to glycoproteins can reveal the points of oligosaccharide attachment. Cleavage by a specific protease yields a characteristic pattern of peptide fragments that can be analyzed chromatographically. Fragments attached to oligosaccharides can be picked out because their chromatographic properties will change on glycosidase treatment. Mass spectrometric analysis or direct peptide sequencing can reveal the identity of the peptide in question and, with additional effort, the exact site of oligosaccharide attachment.

While the sequencing of the human genome is complete, the characterization of the much more complex proteome, including the biological roles of glycosylated proteins, presents a challenge to biochemistry.

11.4 Lectins Are Specific Carbohydrate-Binding Proteins

The diversity and complexity of the carbohydrate units and the variety of ways in which they can be joined in oligosaccharides and polysaccharides suggest that they are functionally important. Nature does not construct complex patterns when simple ones suffice. Why all this intricacy and diversity? It is now clear that these carbohydrate structures are the recognition sites for a special class of proteins. Such proteins, termed *glycan-binding proteins,* bind specific carbohydrate structures on neighboring cell surfaces. Originally discovered in plants, glycan-binding proteins are ubiquitous, and no living organisms have been found that lack these key proteins. We will focus on a particular class of glycan-binding proteins termed *lectins* (from Latin *legere,* "to select"). The interaction of lectins with their carbohydrate partners is another example of carbohydrates being information-rich molecules that guide many biological processes. The diverse carbohydrate structures displayed on cell surfaces are well-suited to serving as sites of interaction between cells and their environments. Interestingly, the partners for lectin binding are often the carbohydrate moiety of glycoproteins.

Lectins promote interactions between cells

Cell–cell contact is a vital interaction in a host of biochemical functions, ranging from building a tissue from isolated cells to facilitating the transmission of information. The chief function of lectins is to facilitate cell–cell contact. A lectin usually contains two or more carbohydrate-binding sites. The lectins on the surface of one cell interact with arrays of carbohydrates displayed on the surface of another cell. Lectins and carbohydrates are linked by a number of weak noncovalent interactions that ensure specificity yet permit unlinking as needed. The weak interactions between one cell surface and another resemble the action of Velcro; each interaction is weak, but the composite is strong.

We have already met a lectin obliquely. Recall that, in I-cell disease, lysosomal enzymes lack the appropriate mannose 6-phosphate, a molecule that directs the enzymes to the lysosome. Under normal circumstances, the *mannose 6-phosphate receptor,* a lectin, binds the enzymes in the Golgi apparatus and directs them to the lysosome.

Lectins are organized into different classes

Lectins can be divided into classes on the basis of their amino acid sequences and biochemical properties. One large class is the C type (for *c*alcium-requiring) found in animals. These proteins each have a homologous domain of 120 amino acids that is responsible for carbohydrate binding. The structure of one such domain bound to a carbohydrate target is shown in Figure 11.30.

A calcium ion on the protein acts as a bridge between the protein and the sugar through direct interactions with sugar OH groups. In addition, two glutamate residues in the protein bind to both the calcium ion and the sugar, and other protein side chains form hydrogen bonds with other

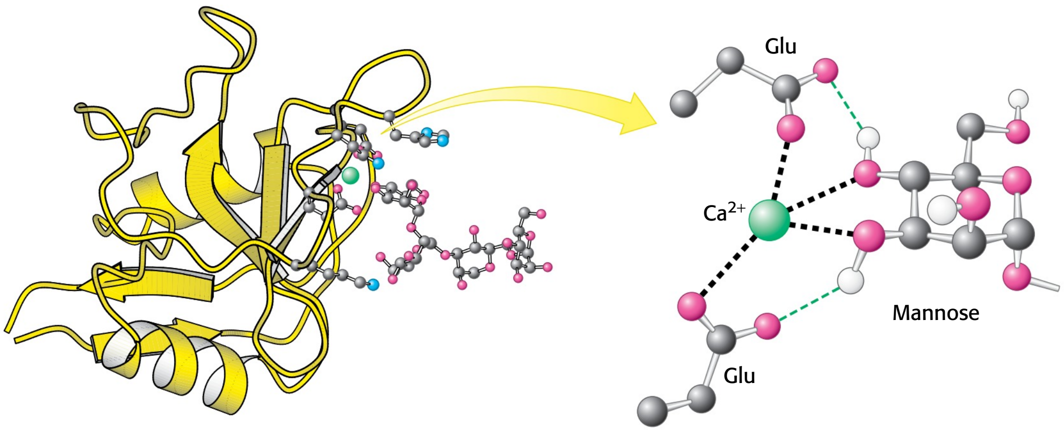


FIGURE 11.30 Structure of a carbohydrate-binding domain of an animal C-type lectin. *Notice* that a calcium ion links a mannose residue to the lectin. Selected interactions are shown, with some hydrogen atoms omitted for clarity. [Drawn from 2MSC. pdb.]

OH groups on the carbohydrate. The carbohydrate-binding specificity of a particular lectin is determined by the amino acid residues that bind the carbohydrate. C-type lectins function in a variety of cellular activities, including receptor-mediated endocytosis, a process by which soluble molecules are bound to the cell surface and subsequently internalized (Section 26.3), and cell–cell recognition.

Proteins termed *selectins* are members of the C-type family. Selectins bind immune-system cells to sites of injury in the inflammatory response. The L, E, and P forms of selectins bind specifically to carbohydrates on *l*ymph-node vessels, *e*ndothelium, or activated blood *p*latelets, respectively. New therapeutic agents that control inflammation may emerge from a deeper understanding of how selectins bind and distinguish different carbohydrates. *L*-Selectin, originally thought to participate only in the immune response, is produced by embryos when they are ready to attach to the endometrium of the mother's uterus. For a short period of time, the endometrial cells present an oligosaccharide on the cell surface. When the embryo attaches through lectins, the attachment activates signal pathways in the endometrium to make implantation of the embryo possible.

Another large class of lectins comprises the *L*-lectins. These lectins are especially rich in the seeds of leguminous plants, and many of the initial biochemical characterizations of lectins were performed on this readily available lectin. Although the exact role of lectins in plants is unclear, they can serve as potent insecticides. Other *L*-type lectins, such as *calnexin* and *calreticulin,* are prominent chaperones in the eukaryotic endoplasmic reticulum. Recall that chaperones are proteins that facilitate the folding of other proteins (p. 51).

Influenza virus binds to sialic acid residues

Many pathogens gain entry into specific host cells by adhering to cell-surface carbohydrates. For example, influenza virus recognizes sialic acid residues linked to galactose residues that are present on cell-surface glycoproteins. The viral protein that binds to these sugars is called *hemagglutinin* (Figure 11.31A).

After binding hemagglutinin, the virus is engulfed by the cell and begins to replicate. To exit the cell, a process essentially the reverse of viral entry occurs (Figure 11.31B). Viral assembly results in the budding of the viral particle from the cell. Upon complete assembly, the viral particle is still attached to sialic acid residues of the cell membrane by hemagglutinin on the surface of the new virions. Another viral protein, neuraminidase (sialidase),

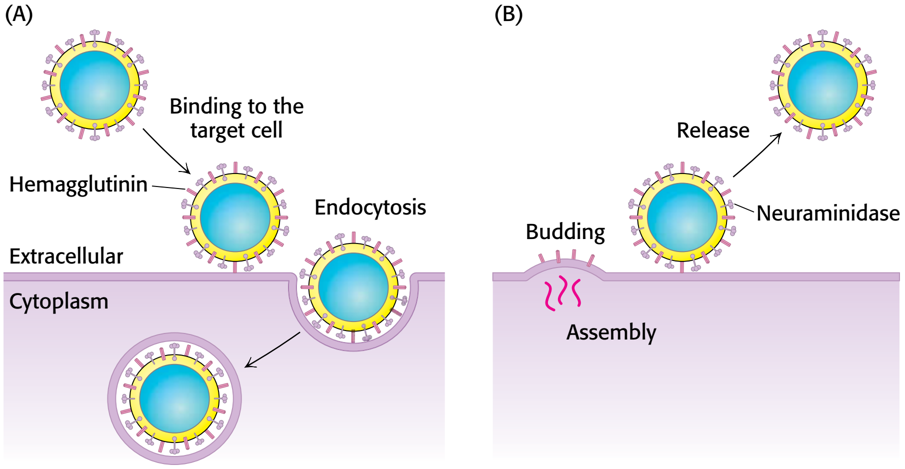


FIGURE 11.31 Viral receptors. (A) Influenza virus targets cells by binding to sialic acid residues located at the termini of oligosaccharides present on cell-surface glycoproteins and glycolipids. These carbohydrates are bound by hemagglutinin, one of the major proteins expressed on the surface of the virus. (B) When viral replication is complete and the viral particle buds from the cell, the other major viral-surface protein, neuraminidase, cleaves oligosaccharide chains to release the viral particle.

cleaves the glycosidic bonds between the sialic acid residues and the rest of the cellular glycoprotein, freeing the virus to infect new cells, and thus spreading the infection throughout the respiratory tract. Inhibitors of this enzyme such as oseltamivir (Tamiflu) and zanamivir (Relenza) are important anti-influenza agents.

Viral hemagglutinin's carbohydrate-binding specificity may play an important role in species specificity of infection and ease of transmission. For instance, avian influenza H5N1 (bird flu) is especially lethal and is readily spread from bird to bird. Although human beings can be infected by this virus, infection is rare and human-to-human transmission is rarer still. The biochemical basis of these characteristics is that the avian-virus hemagglutinin recognizes a different carbohydrate sequence from that recognized in human influenza. Although human beings have the sequence to which the avian virus binds, it is located deep in the lungs. Infection by the avian virus is thus difficult, and, when it does occur, the avian virus is not readily transmitted by sneezing or coughing.

Plasmodium falciparum, the parasitic protozoan that causes malaria, also relies on glycan binding to infect and colonize its host. Glycan-binding proteins of the parasitic form initially injected by the mosquito bind to the glycosaminoglycan heparin sulfate on the liver, initiating the parasite's entry into the cell. On exiting from the liver later in its life cycle, the parasite invades red blood cells by using another glycan-binding protein to bind to the carbohydrate moiety of *glycophorin*, a prominent membrane glycoprotein in red blood cells. Developing means to disrupt the carbohydrate interactions between pathogens and host cells may prove to be clinically useful.

SUMMARY

11.1 Monosaccharides Are the Simplest Carbohydrates

Carbohydrates are aldoses or ketoses that are rich in hydroxyl groups. An aldose is a carbohydrate with an aldehyde group (as in glyceraldehyde and glucose), whereas a ketose contains a keto group (as in dihydroxyacetone and fructose). A sugar belongs to the D series if the absolute configuration of its asymmetric carbon atom farthest from the aldehyde or keto group is the same as that of D-glyceraldehyde. Most naturally occurring sugars belong to the D series. The C-1 aldehyde in the open-chain form of glucose reacts with the C-5 hydroxyl group to form a six-membered pyranose ring. The C-2 keto group in the open-chain form of fructose reacts with the C-5 hydroxyl group to form a five-membered furanose ring. Pentoses such as ribose and deoxyribose also form furanose rings. An additional asymmetric center is formed at the anomeric carbon atom (C-1 in aldoses and C-2 in ketoses) in these cyclizations. The hydroxyl group attached to the anomeric carbon atom is on the opposite side of the ring from the CH_2OH group attached to the chiral center in the α anomer, whereas it is on the same side of the ring as the CH_2OH group in the β anomer. Not all atoms in the ring lie in the same plane. Rather, pyranose rings usually adopt the chair conformation, and furanose rings usually adopt the envelope conformation. Sugars are joined to alcohols and amines by glycosidic bonds from the anomeric carbon atom. For example, *N*-glycosidic bonds link sugars to purines and pyrimidines in nucleotides, RNA, and DNA.

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates

Sugars are linked to one another in disaccharides and polysaccharides by *O*-glycosidic bonds. Sucrose, lactose, and maltose are the common disaccharides. Sucrose (common table sugar) consists of α-glucose and β-fructose joined by a glycosidic linkage between their anomeric carbon atoms. Lactose (in milk) consists of galactose joined to glucose by a β-1,4 linkage. Maltose (from starch) consists of two glucoses joined by an α-1,4 linkage. Starch is a polymeric form of glucose in plants, and glycogen serves a similar role in animals. Most of the glucose units in starch and glycogen are in α-1,4 linkage. Cellulose, the major structural polymer of plant cell walls, consists of glucose units joined by β-1,4 linkages. These β linkages give rise to long straight chains that form fibrils with high tensile strength. In contrast, the α linkages in starch and glycogen lead to open helices, in keeping with their roles as mobilizable energy stores.

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins

Carbohydrates are commonly conjugated to proteins. If the protein component is predominant, the conjugate of protein and carbohydrate is called a glycoprotein. Most secreted proteins, such as the signal molecule erythropoietin, are glycoproteins. Glycoproteins are also prominent on the external surface of the plasma membrane. Proteins bearing covalently linked glycosaminoglycans are proteoglycans. Glycosaminoglycans are polymers of repeating disaccharides. One of the units in each repeat is a derivative of glucosamine or galactosamine. These highly anionic carbohydrates have a high density of carboxylate or sulfate groups. Proteoglycans are found in the extracellular matrices of animals and are key components of cartilage. Mucoproteins, like proteoglycans, are predominantly carbohydrate by weight. The protein component is heavily *O*-glycosylated with *N*-acetylgalactosamine joining the oligosaccharide to the protein. Mucoproteins serve as lubricants.

Glycosyltransferases link the oligosaccharide units on proteins either to the side-chain oxygen atom of a serine or threonine residue or to the side-chain amide nitrogen atom of an asparagine residue. Protein glycosylation takes place in the lumen of the endoplasmic reticulum. The *N*-linked oligosaccharides are synthesized on dolichol phosphate and subsequently transferred to the protein acceptor. Additional sugars are attached in the Golgi complex to form diverse patterns.

11.4 Lectins Are Specific Carbohydrate-Binding Proteins

Carbohydrates on cell surfaces are recognized by proteins called lectins. In animals, the interplay of lectins and their sugar targets guides cell–cell contact. The viral protein hemagglutinin on the surface of the influenza virus recognizes sialic acid residues on the surfaces of cells invaded by the virus.

KEY TERMS

glycobiology (p. 316)
glycomics (p. 316)
monosaccharide (p. 316)
ketose (p. 316)
aldose (p. 316)
constitutional isomer (p. 317)
stereoisomer (p. 317)
enantiomer (p. 317)
diastereoisomer (p. 317)
epimer (p. 318)
hemiacetal (p. 318)
pyranose (p. 318)
hemiketal (p. 318)
furanose (p. 319)
anomer (p. 319)
reducing sugar (p. 321)
nonreducing sugar (p. 321)
advanced glycation end product (AGE) (p. 322)
glycosidic bond (p. 322)
oligosaccharide (p. 323)
disaccharide (p. 323)
polysaccharide (p. 324)
glycogen (p. 324)

starch (p. 324)
cellulose (p. 324)
glycoprotein (p. 325)
proteoglycan (p. 326)
glycosaminoglycan (p. 326)
glycoform (p. 326)
mucin (mucoprotein) (p. 329)
endoplasmic reticulum (p. 330)
Golgi complex (p. 330)
dolichol phosphate (p. 330)
glycosyltransferase (p. 331)
glycan-binding protein (p. 334)
lectin (p. 334)
selectin (p. 335)

PROBLEMS

1. *Word origin*. Account for the origin of the term *carbohydrate*.

2. *Diversity*. How many different oligosaccharides can be made by linking one glucose, one mannose, and one galactose? Assume that each sugar is in its pyranose form. Compare this number with the number of tripeptides that can be made from three different amino acids.

3. *They go together like a horse and carriage*. Match each term with its description.

(a) Enantiomers _____
(b) Cellulose _____
(c) Lectins _____
(d) Glycosyltransferases _____
(e) Epimers _____
(f) Starch _____
(g) Carbohydrates _____
(h) Proteoglycan _____
(i) Mucoprotein _____
(j) Glycogen _____

1. Has the molecular formula of $(CH_2O)_n$
2. Monosaccharides that differ at a single asymmetric carbon atom
3. The storage form of glucose in animals
4. The storage form of glucose in plants
5. Glycoprotein containing glycosaminoglycans
6. The most abundant organic molecule in the biosphere
7. *N*-Acetylgalactosamine is a key component of this glycoprotein
8. Carbohydrate-binding proteins
9. Enzymes that synthesize oligosaccharides
10. Stereoisomers that are mirror images of each other

4. *Couples*. Indicate whether each of the following pairs of sugars consists of anomers, epimers, or an aldose–ketose pair:

(a) D-glyceraldehyde and dihydroxyacetone

(b) D-glucose and D-mannose

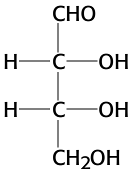
(c) D-glucose and D-fructose

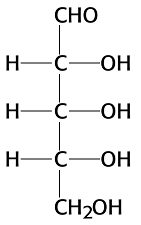
(d) α-D-glucose and β-D-glucose

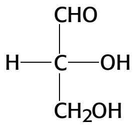
(e) D-ribose and D-ribulose

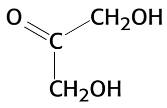
(f) D-galactose and D-glucose

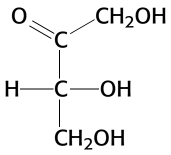
5. *Carbons and carbonyls*. To which classes of sugars do the monosaccharides shown here belong?

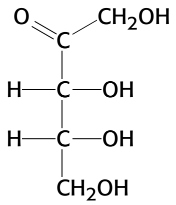

D-Erythrose

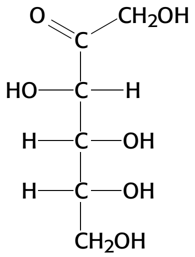

D-Ribose


D-Glyceraldehyde


Dihydroxyacetone

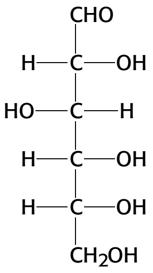

D-Erythrulose


D-Ribulose


D-Fructose

6. *Chemical cousins*. Although an aldose with 4 asymmetric carbon atoms is capable of forming 16 diastereoisomers, only 8 of the isomers are commonly observed, including glucose. They are listed below with their structural relation to glucose. Using the structure of glucose as a reference, draw the structures.

(a) D-Allose: Epimeric at C-3
(b) D-Altrose: Isomeric at C-2 and C-3
(c) D-Mannose: Epimeric at C-2
(d) D-Gulose: Isomeric at C-3 and C-4
(e) D-Idose: Isomeric at C-2, C-3 and C-4
(f) D-Galactose: Epimeric at C-4
(g) D-Talose: Isomeric at C-2 and C-4


D-Glucose

7. *An art project*. Draw the structure of the disaccharide α-glycosyl-(1→6)-galactose in the β anomeric form.

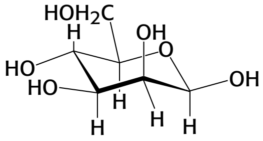
8. *Mutarotation.* The specific rotations of the α and β anomers of D-glucose are +112 degrees and +18.7 degrees, respectively. Specific rotation, $[\alpha]_D$, is defined as the observed rotation of light of wavelength 589 nm (the D line of a sodium lamp) passing through 10 cm of a 1 g ml^{-1} solution of a sample. When a crystalline sample of α-D-glucose is dissolved in water, the specific rotation decreases from 112 degrees to an equilibrium value of 52.7 degrees. On the basis of this result, what are the proportions of the α and β anomers at equilibrium? Assume that the concentration of the open-chain form is negligible.

9. *Telltale marker.* Glucose reacts slowly with hemoglobin and other proteins to form covalent compounds. Why is glucose reactive? What is the nature of the adduct formed?

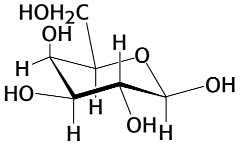
10. *Periodate cleavage.* Compounds containing hydroxyl groups on adjacent carbon atoms undergo carbon–carbon bond cleavage when treated with periodate ion (IO_4^-). How can this reaction be used to distinguish between pyranosides and furanosides?

11. *Sugar lineup.* Identify the following four sugars.

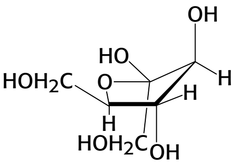
(a)



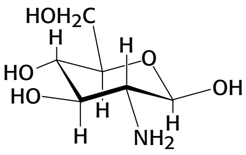
(b)



(c)



(d)



12. *Cellular glue.* A trisaccharide unit of a cell-surface glycoprotein is postulated to play a critical role in mediating cell–cell adhesion in a particular tissue. Design a simple experiment to test this hypothesis.

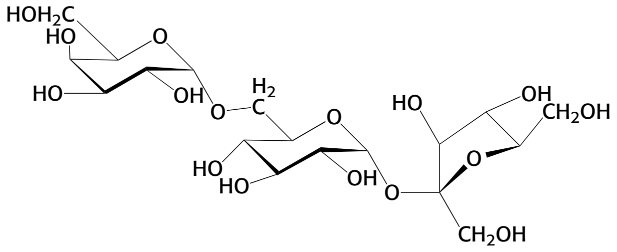
13. *Mapping the molecule.* Each of the hydroxyl groups of glucose can be methylated with reagents such as dimethyl sulfate under basic conditions. Explain how exhaustive methylation followed by the complete digestion of a known amount of glycogen would enable you to determine the number of branch points and reducing ends.

14. *Component parts.* Raffinose is a trisaccharide and a minor constituent in sugar beets.

(a) Is raffinose a reducing sugar? Explain.

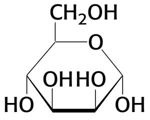
(b) What are the monosaccharides that compose raffinose?

(c) β-Galactosidase is an enzyme that will remove galactose residues from an oligosaccharide. What are the products of β-galactosidase treatment of raffinose?



Raffinose

15. *Anomeric differences.* α-D-Mannose is a sweet-tasting sugar. β-D-Mannose, on the other hand, tastes bitter. A pure solution of α-D-mannose loses its sweet taste with time as it is converted into the β anomer. Draw the β anomer and explain how it is formed from the α anomer.



α-D-Mannose

16. *A taste of honey.* Fructose in its β-D-pyranose form accounts for the powerful sweetness of honey. The β-D-furanose form, although sweet, is not as sweet as the pyranose form. The furanose form is the more stable form. Draw the two forms and explain why it may not always be wise to cook with honey.

17. *Making ends meet.*

(a) Compare the number of reducing ends to nonreducing ends in a molecule of glycogen.

(b) As we will see in Chapter 21, glycogen is an important fuel-storage form that is rapidly mobilized. At which end—the reducing or nonreducing—would you expect most metabolism to take place?

18. *A lost property*. Glucose and fructose are reducing sugars. Sucrose, or table sugar, is a disaccharide consisting of both fructose and glucose. Is sucrose a reducing sugar? Explain.

19. *Meat and potatoes*. Compare the structures of glycogen and starch.

20. *Straight or with a twist?* Account for the different structures of glycogen and cellulose.

21. *Sweet proteins*. List the key classes of glycoproteins, their defining characteristics, and their biological functions.

22. *Life extender*. What is the function of the carbohydrate moiety that is attached to EPO?

23. *Cushioning*. What is the role of the glycosaminoglycan in the cushioning provided by cartilage?

24. *Undelivered mail. Not returned to sender*. I-cell disease results when proteins normally destined to the lysosomes lack the appropriate carbohydrate-addressing molecule (p. 332). Suggest another possible means by which I-cell disease might arise.

25. *Appropriate peg.* Which amino acids are used for the attachment of carbohydrates to proteins?

26. *From one, many.* What is meant by a glycoform?

27. *Ome.* What is meant by the glycome?

28. *Exponential expansion?* Compare the amount of information inherent in the genome, the proteome, and the glycome.

29. *Attachments.* Suppose, one Sunday afternoon, you are relaxing by reading amino acid sequences of various proteins. Being a bit hungry, you are also thinking of sweet snacks. Combining your interests, you wonder whether you can detect *N*-glycosylation sites by simply looking at amino acid sequence. Your roomate, who is taking a biochemistry course, says "You sure can, to some degree at least, and here's how." What did your roomate say to explain?

30. *Locks and keys.* What does the fact that all organisms contain lectins suggest about the role of carbohydrates?

31. *Carbohydrates—not just for breakfast anymore.* Differentiate between a glycoprotein and a lectin.

32. *Carbohydrates and proteomics.* Suppose that a protein contains six potential *N*-linked glycosylation sites. How many possible proteins can be generated, depending on which of these sites is actually glycosylated? Do not include the effects of diversity within the carbohydrate added.

Chapter Integration Problems

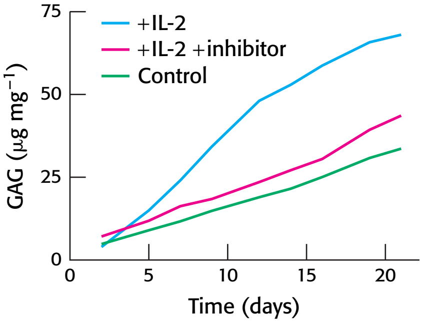
33. *Like a jigsaw puzzle.* Why is it more difficult to determine the structure of the oligosaccharides compared to amino acid sequences or nucleotide sequences?

34. *Stereospecificity.* Sucrose, a major product of photosynthesis in green leaves, is synthesized by a battery of enzymes. The substrates for sucrose synthesis, D-glucose and D-fructose, are a mixture of α and β anomers as well as acyclic compounds in solution. Nonetheless, sucrose consists of α-D-glucose linked by its carbon-1 atom to the carbon-2 atom of β-D-fructose. How can the specificity of sucrose be explained in light of the potential substrates?

35. *Specific recognition.* How might the technique of affinity chromatography be used to purify lectins?

Data Interpretation Problem

36. *Sore joints.* A contributing factor to the development of arthritis is the inappropriate proteolytic destruction of the aggrecan component of cartilage by the proteolytic enzyme aggrecanase. The immune-system signal molecule interleukin 2 (IL-2) activates aggrecanase; in fact, IL-2 blockers are sometimes used to treat arthritis. Studies were undertaken to determine whether inhibitors of aggrecanase can counteract the effects of IL-2. Pieces of cartilage were incubated in media with various additions and the amount of aggrecan destruction was measured as a function of time.



[Data from M.A. Pratta et al., *J. Biol. Chem.* 278:45539–45545, 2003, Fig. 7B.]

(a) Aggrecan degradation was measured by the release of glycosaminoglycan. What is the rationale for this assay?

(b) Why might glycosaminoglycan release not indicate aggrecan degradation?

(c) What is the purpose of the control—cartilage incubated with no additions?

(d) What is the effect of adding IL-2 to the system?

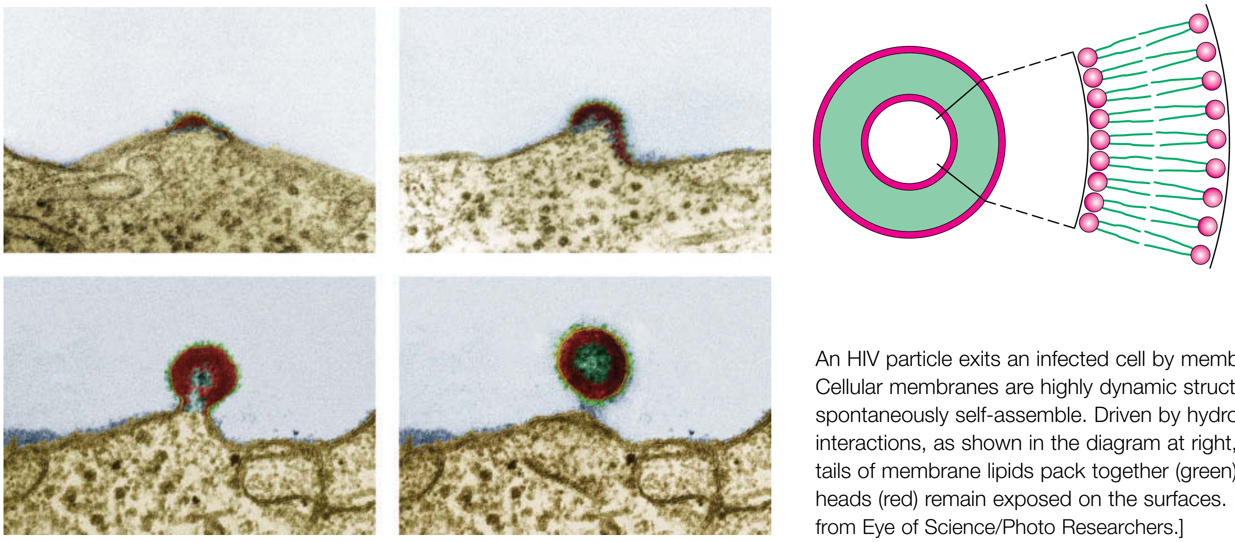
(e) What is the response when an aggrecanase inhibitor is added in addition to IL-2?

(f) Why is there some aggrecan destruction in the control with the passage of time?

CHAPTER

12

Lipids and Cell Membranes



An HIV particle exits an infected cell by membrane budding. Cellular membranes are highly dynamic structures that spontaneously self-assemble. Driven by hydrophobic interactions, as shown in the diagram at right, the fatty acid tails of membrane lipids pack together (green), while the polar heads (red) remain exposed on the surfaces. [Micrographs from Eye of Science/Photo Researchers.]

The boundaries of all cells are defined by *biological membranes* (Figure 12.1), dynamic structures in which proteins float in a sea of lipids. The lipid component prevents molecules generated inside the cell from leaking out and unwanted molecules from diffusing in, while the protein components act as transport systems that allow the cell to take up specific molecules and remove unwanted ones. Such transport systems confer on membranes the important property of *selective permeability*. We will consider these transport systems in greater detail in the next chapter.

In addition to an external cell membrane (called the *plasma membrane*), eukaryotic cells also contain internal membranes that form the boundaries of organelles such as mitochondria, chloroplasts, peroxisomes, and lysosomes. Functional specialization in the course of evolution has been closely linked to the formation of such compartments. Specific systems have evolved to allow the targeting of selected proteins into or through particular internal membranes and, hence, into specific organelles. External and internal membranes share essential properties; these features are the subject of this chapter.

Biological membranes serve several additional functions indispensable for life, such as energy storage and information transduction, that are dictated by the proteins associated with them. In this chapter, we will examine the properties of membrane proteins that enable them to exist in the hydrophobic environment of the membrane while connecting two hydrophilic environments. In the next chapter, we will discuss the functions of these proteins.

OUTLINE

12.1 Fatty Acids Are Key Constituents of Lipids

12.2 There Are Three Common Types of Membrane Lipids

12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media

12.4 Proteins Carry Out Most Membrane Processes

12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane

12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes

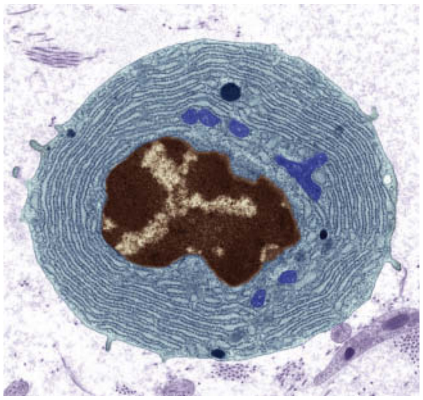


FIGURE 12.1 Electron micrograph of a plasma cell. This image has been colored to indicate the distinct boundary of the cell, formed by its plasma membrane. [Steve Gschmeissner/Photo Researchers.]

Many Common Features Underlie the Diversity of Biological Membranes

Membranes are as diverse in structure as they are in function. However, they do have in common a number of important attributes:

1. Membranes are *sheetlike structures,* only two molecules thick, that form *closed boundaries* between different compartments. The thickness of most membranes is between 60 Å (6 nm) and 100 Å (10 nm).

2. Membranes consist mainly of *lipids* and *proteins*. The mass ratio of lipids to proteins ranges from 1:4 to 4:1. Membranes also contain *carbohydrates* that are linked to lipids and proteins.

3. Membrane lipids are small molecules that have both *hydrophilic* and *hydrophobic* moieties. These lipids spontaneously form *closed bimolecular sheets* in aqueous media. These *lipid bilayers* are barriers to the flow of polar molecules.

4. *Specific proteins mediate distinctive functions of membranes*. Proteins serve as pumps, channels, receptors, energy transducers, and enzymes. Membrane proteins are embedded in lipid bilayers, which create suitable environments for their action.

5. Membranes are *noncovalent assemblies*. The constituent protein and lipid molecules are held together by many noncovalent interactions, which act cooperatively.

6. Membranes are *asymmetric*. The two faces of biological membranes always differ from each other.

7. Membranes are *fluid structures*. Lipid molecules diffuse rapidly in the plane of the membrane, as do proteins, unless they are anchored by specific interactions. In contrast, lipid molecules and proteins do not readily rotate across the membrane. Membranes can be regarded as *two-dimensional solutions of oriented proteins and lipids*.

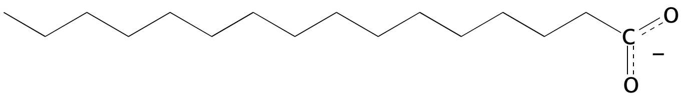
8. Most cell membranes are *electrically polarized,* such that the inside is negative [typically −60 millivolts (mV)]. Membrane potential plays a key role in transport, energy conversion, and excitability (Chapter 13).

12.1 Fatty Acids Are Key Constituents of Lipids

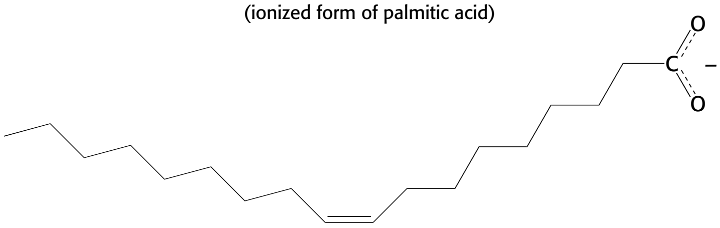
The hydrophobic properties of lipids are essential to their ability to form membranes. Most lipids owe their hydrophobic properties to one component, their fatty acids.

Fatty acid names are based on their parent hydrocarbons

Fatty acids are long hydrocarbon chains of various lengths and degrees of unsaturation that terminate with carboxylic acid groups. The systematic name for a fatty acid is derived from the name of its parent hydrocarbon by the substitution of *oic* for the final *e*. For example, the C_{18} saturated fatty acid is called *octadecanoic acid* because the parent hydrocarbon is octadecane. A C_{18} fatty acid with one double bond is called octadec*enoic* acid; with two double bonds, octadeca*dienoic* acid; and with three double bonds, octadeca*trienoic* acid. The notation 18:0 denotes a C_{18} fatty acid with no double bonds, whereas 18:2 signifies that there are two double bonds. The structures of the ionized forms of two common fatty acids—palmitic acid (16:0) and oleic acid (18:1)—are shown in Figure 12.2.



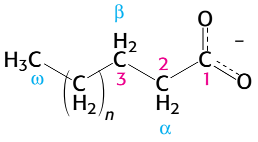
Palmitate
(ionized form of palmitic acid)

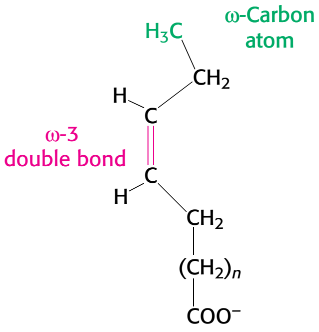


Oleate
(ionized form of oleic acid)

FIGURE 12.2 Structures of two fatty acids. Palmitate is a 16-carbon, saturated fatty acid, and oleate is an 18-carbon fatty acid with a single cis double bond.

Fatty acid carbon atoms are numbered starting at the carboxyl terminus, as shown in the margin. Carbon atoms 2 and 3 are often referred to as α and β, respectively. The methyl carbon atom at the distal end of the chain is called the *ω-carbon atom.* The position of a double bond is represented by the symbol Δ followed by a superscript number. For example, *cis*-Δ^9 means that there is a cis double bond between carbon atoms 9 and 10; *trans*-Δ^2 means that there is a trans double bond between carbon atoms 2 and 3. Alternatively, the position of a double bond can be denoted by counting from the distal end, with the ω-carbon atom (the methyl carbon) as number 1. An ω-3 fatty acid, for example, has the structure shown in the margin. Fatty acids are ionized at physiological pH, and so it is appropriate to refer to them according to their carboxylate form: for example, palmitate or hexadecanoate.





An ω-3 fatty acid

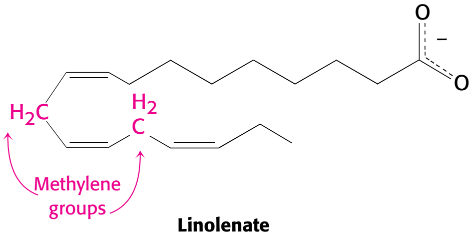
Fatty acids vary in chain length and degree of unsaturation

Fatty acids in biological systems usually contain an even number of carbon atoms, typically between 14 and 24 (Table 12.1). The 16- and 18-carbon fatty acids are most common. The dominance of fatty acid chains containing an even number of carbon atoms reflects the manner in which fatty acids are biosynthesized (Chapter 26). The hydrocarbon chain is almost invariably unbranched in animal fatty acids. The alkyl chain may be saturated or it may contain one or more double bonds. The configuration of the double bonds in most unsaturated fatty acids is cis. The double

TABLE 12.1 Some naturally occurring fatty acids in animals

Number of carbons	Number of double bonds	Common name	Systematic name	Formula
12	0	Laurate	*n*-Dodecanoate	$CH_3(CH_2)_{10}COO^-$
14	0	Myristate	*n*-Tetradecanoate	$CH_3(CH_2)_{12}COO^-$
16	0	Palmitate	*n*-Hexadecanoate	$CH_3(CH_2)_{14}COO^-$
18	0	Stearate	*n*-Octadecanoate	$CH_3(CH_2)_{16}COO^-$
20	0	Arachidate	*n*-Eicosanoate	$CH_3(CH_2)_{18}COO^-$
22	0	Behenate	*n*-Docosanoate	$CH_3(CH_2)_{20}COO^-$
24	0	Lignocerate	*n*-Tetracosanoate	$CH_3(CH_2)_{22}COO^-$
16	1	Palmitoleate	*cis*-Δ^9-Hexadecenoate	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COO^-$
18	1	Oleate	*cis*- Δ^9-Octadecenoate	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COO^-$
18	2	Linoleate	*cis,cis*- Δ^9, Δ^{12}-Octadecadienoate	$CH_3(CH_2)_4(CH{=}CHCH_2)_2(CH)_6COO^-$
18	3	Linolenate	all-*cis*- Δ^9, Δ^{12}, Δ^{15}-Octadecatrienoate	$CH_3CH_2(CH{=}CHCH_2)_3(CH_2)_6COO^-$
20	4	Arachidonate	all-*cis* Δ^5, Δ^8, Δ^{11}, Δ^{14}-Eicosatetraenoate	$CH_3(CH2)_4(CH{=}CHCH_2)_4(CH_2)_2COO^-$

bonds in polyunsaturated fatty acids are separated by at least one methylene group.



Linolenate

The properties of fatty acids and of lipids derived from them are markedly dependent on chain length and degree of saturation. Unsaturated fatty acids have lower melting points than do saturated fatty acids of the same length. For example, the melting point of stearic acid is 69.6°C, whereas that of oleic acid (which contains one cis double bond) is 13.4°C. The melting points of polyunsaturated fatty acids of the C_{18} series are even lower. Chain length also affects the melting point, as illustrated by the fact that the melting temperature of palmitic acid (C_{16}) is 6.5 degrees lower than that of stearic acid (C_{18}). Thus, *short chain length and unsaturation enhance the fluidity of fatty acids and of their derivatives.*

12.2 There Are Three Common Types of Membrane Lipids

By definition, *lipids are water-insoluble biomolecules that are highly soluble in organic solvents such as chloroform.* Lipids have a variety of biological roles: they serve as fuel molecules, highly concentrated energy stores, signal molecules and messengers in signal-transduction pathways, and components of membranes. The first three roles of lipids will be considered in later chapters. Here, our focus is on lipids as membrane constituents. The three major kinds of membrane lipids are *phospholipids, glycolipids,* and *cholesterol.* We begin with lipids found in eukaryotes and bacteria. The lipids in archaea are distinct, although they have many features related to membrane formation in common with lipids of other organisms.

Phospholipids are the major class of membrane lipids

Phospholipids are abundant in all biological membranes. A phospholipid molecule is constructed from four components: one or more fatty acids, a platform to which the fatty acids are attached, a phosphate, and an alcohol attached to the phosphate (Figure 12.3). The fatty acid components provide a hydrophobic barrier, whereas the remainder of the molecule has hydrophilic properties that enable interaction with the aqueous environment.

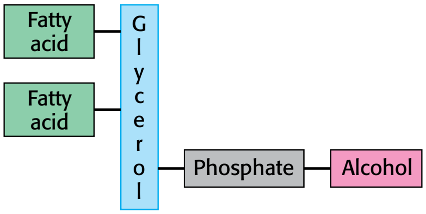
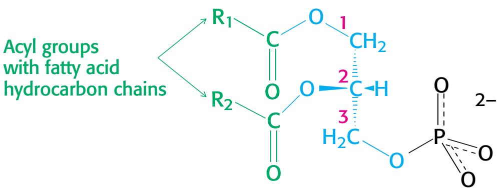


FIGURE 12.3 Schematic structure of a phospholipid.

The platform on which phospholipids are built may be *glycerol,* a three-carbon alcohol, or *sphingosine,* a more complex alcohol. Phospholipids derived from glycerol are called *phosphoglycerides.* A phosphoglyceride consists of a glycerol backbone to which are attached two fatty acid chains and a phosphorylated alcohol.

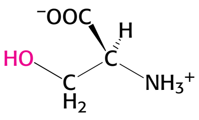
In phosphoglycerides, the hydroxyl groups at C-1 and C-2 of glycerol are esterified to the carboxyl groups of the two fatty acid chains. The C-3 hydroxyl group of the glycerol backbone is esterified to phosphoric acid. When no further additions are made, the resulting compound is *phosphatidate (diacylglycerol 3-phosphate),* the simplest phosphoglyceride. Only small amounts of phosphatidate are present in membranes. However, the molecule is a key intermediate in the biosynthesis of the other phosphoglycerides (Section 26.1). The absolute configuration of the glycerol 3-phosphate moiety of membrane lipids is shown in Figure 12.4.



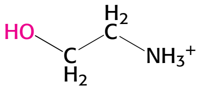
Phosphatidate (Diacylglycerol 3-phosphate)

FIGURE 12.4 Structure of phosphatidate (diacylglycerol 3-phosphate). The absolute configuration of the center carbon (C-2) is shown.

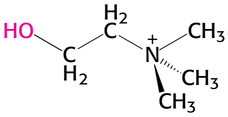
The major phosphoglycerides are derived from phosphatidate by the formation of an ester bond between the phosphate group of phosphatidate and the hydroxyl group of one of several alcohols. The common alcohol moieties of phosphoglycerides are the amino acid serine, ethanolamine, choline, glycerol, and inositol.



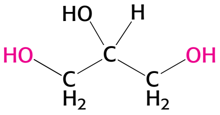
Serine



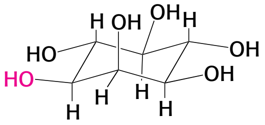
Ethanolamine



Choline

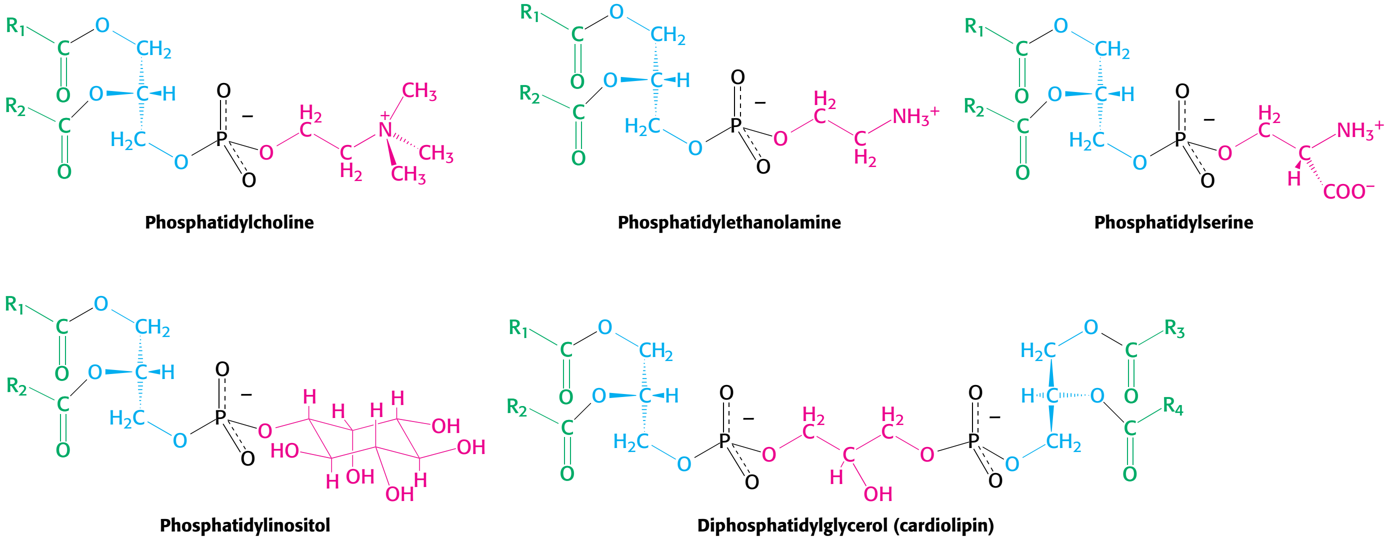


Glycerol



Inositol

The structural formulas of phosphatidylcholine and the other principal phosphoglycerides—namely, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, and diphosphatidylglycerol—are given in Figure 12.5.



Phosphatidylcholine

Phosphatidylethanolamine

Phosphatidylserine

Phosphatidylinositol

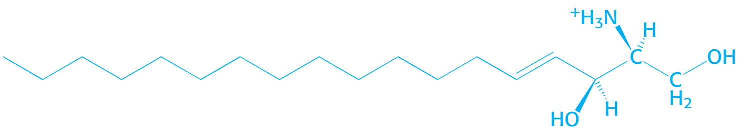
Diphosphatidylglycerol (cardiolipin)

FIGURE 12.5 **Some common phosphoglycerides found in membranes.**

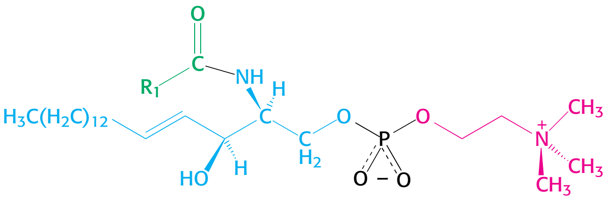
Sphingomyelin is a phospholipid found in membranes that is not derived from glycerol. Instead, the backbone in sphingomyelin is *sphingosine,* an amino alcohol that contains a long, unsaturated hydrocarbon chain (Figure 12.6). In sphingomyelin, the amino group of the sphingosine backbone is linked to a fatty acid by an amide bond. In addition, the primary hydroxyl group of sphingosine is esterified to phosphorylcholine.

Membrane lipids can include carbohydrate moieties

The second major class of membrane lipids, *glycolipids,* are *sugar-containing lipids.* Like sphingomyelin, the glycolipids in animal cells are derived



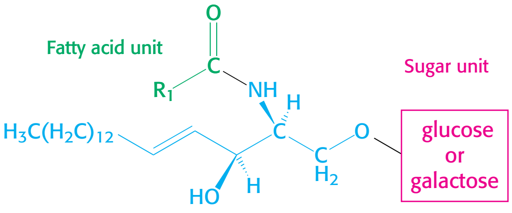
Sphingosine



Sphingomyelin

FIGURE 12.6 **Structures of sphingosine and sphingomyelin.** The sphingosine moiety of sphingomyelin is highlighted in blue.

from sphingosine. The amino group of the sphingosine backbone is acylated by a fatty acid, as in sphingomyelin. Glycolipids differ from sphingomyelin in the identity of the unit that is linked to the primary hydroxyl group of the sphingosine backbone. In glycolipids, one or more sugars (rather than phosphorylcholine) are attached to this group. The simplest glycolipid, called a *cerebroside,* contains a single sugar residue, either glucose or galactose.

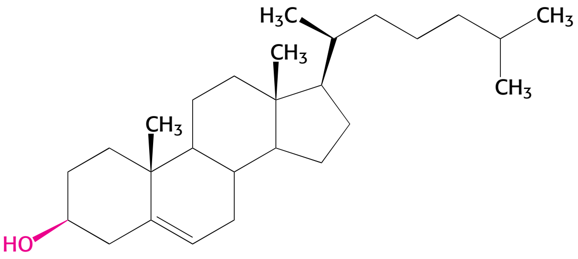


Cerebroside
(a glycolipid)

More-complex glycolipids, such as *gangliosides,* contain a branched chain of as many as seven sugar residues. Glycolipids are oriented in a completely asymmetric fashion with the *sugar residues always on the extracellular side of the membrane.*

Cholesterol Is a Lipid Based on a Steroid Nucleus

Cholesterol, the third major type of membrane lipid, has a structure that is quite different from that of phospholipids. It is a steroid, built from four linked hydrocarbon rings.



Cholesterol

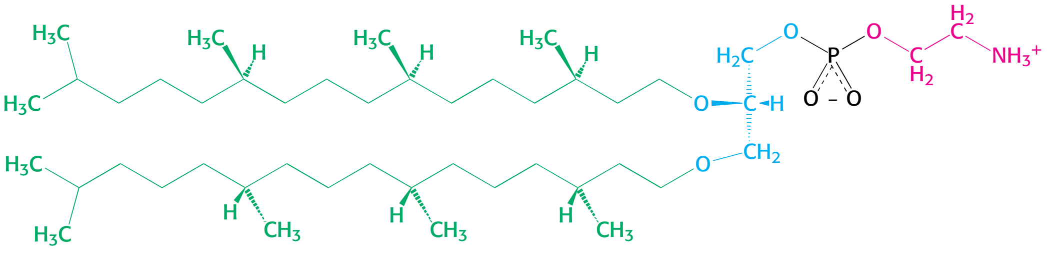
A hydrocarbon tail is linked to the steroid at one end, and a hydroxyl group is attached at the other end. In membranes, the orientation of the molecule is parallel to the fatty acid chains of the phospholipids, and the hydroxyl group interacts with the nearby phospholipid head groups. Cholesterol is absent from prokaryotes but is found to varying degrees in virtually all animal membranes. It constitutes almost 25% of the membrane lipids in certain nerve cells but is essentially absent from some intracellular membranes.

Archaeal membranes are built from ether lipids with branched chains

The membranes of archaea differ in composition from those of eukaryotes or bacteria in three important ways. Two of these differences clearly relate to the hostile living conditions of many archaea (Figure 12.7). First, the nonpolar chains are joined to a glycerol backbone by ether rather than ester linkages. The ether linkage is more resistant to hydrolysis. Second, the alkyl chains are branched rather than linear. They are built up from repeats of a fully saturated five-carbon fragment. These

FIGURE 12.7 An archaeon and its environment. Archaea can thrive in habitats as harsh as a volcanic vent. Here, the archaea form an orange mat surrounded by yellow sulfurous deposits. [Krafft-Explorer/Photo Researchers.]

branched, saturated hydrocarbons are more resistant to oxidation than the unbranched chains of eukaryotic and bacterial membrane lipids. The ability of archaeal lipids to resist hydrolysis and oxidation may help these organisms to withstand the extreme conditions, such as high temperature, low pH, or high salt concentration, under which some of these archaea grow. Finally, the stereochemistry of the central glycerol is inverted compared with that shown in Figure 12.4.

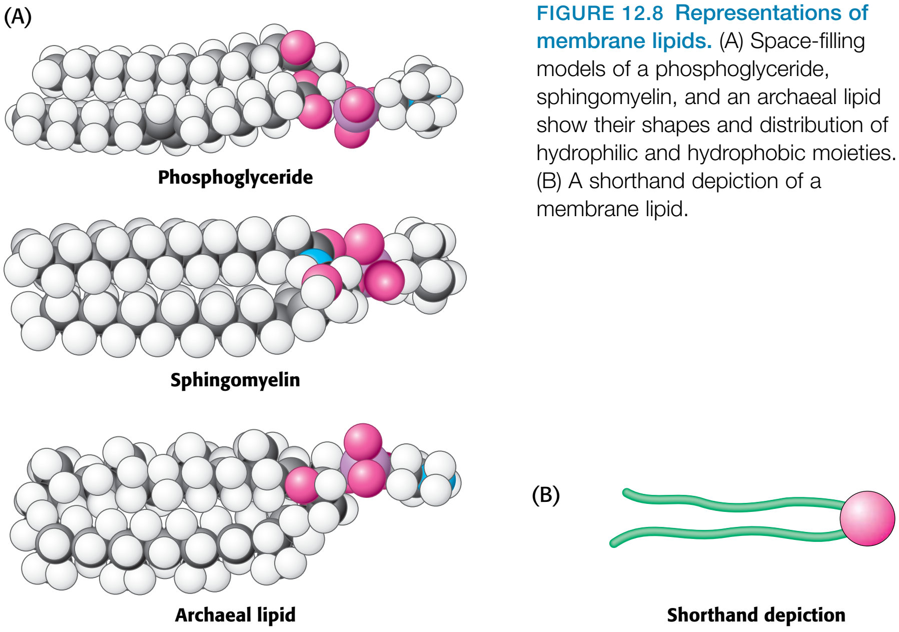


Membrane lipid from the archaeon *Methanococcus jannaschii*

A membrane lipid is an amphipathic molecule containing a hydrophilic and a hydrophobic moiety

The repertoire of membrane lipids is extensive. However, these lipids possess a critical common structural theme: *membrane lipids are amphipathic molecules* (amphiphilic molecules), that is, they contain both a *hydrophilic* and a *hydrophobic* moiety.

Let us look at a model of a phosphoglyceride, such as phosphatidylcholine. Its overall shape is roughly rectangular (Figure 12.8A). The two hydrophobic fatty acid chains are approximately parallel to each other, whereas the hydrophilic phosphorylcholine moiety points in the opposite direction. Sphingomyelin has a similar conformation, as does the archaeal lipid depicted. Therefore, the following shorthand has been adopted to represent these membrane lipids: the hydrophilic unit, also called the *polar head group,* is represented by a circle, and the hydrocarbon tails are depicted by straight or wavy lines (Figure 12.8B).



Phosphoglyceride

Sphingomyelin

Archaeal lipid

Shorthand depiction

FIGURE 12.8 Representations of membrane lipids. (A) Space-filling models of a phosphoglyceride, sphingomyelin, and an archaeal lipid show their shapes and distribution of hydrophilic and hydrophobic moieties. (B) A shorthand depiction of a membrane lipid.

12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media

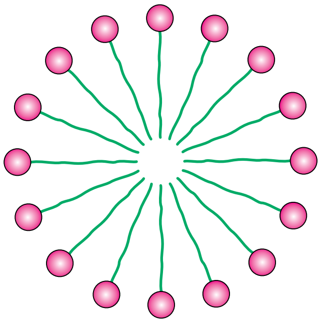


FIGURE 12.9 Diagram of a section of a micelle. Ionized fatty acids readily form such structures, but most phospholipids do not.

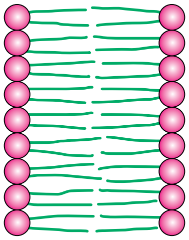


FIGURE 12.10 Diagram of a section of a bilayer membrane.

What properties enable phospholipids to form membranes? *Membrane formation is a consequence of the amphipathic nature of the molecules.* Their polar head groups favor contact with water, whereas their hydrocarbon tails interact with one another in preference to water. How can molecules with these preferences arrange themselves in aqueous solutions? One way is to form a globular structure called a *micelle*. The polar head groups form the outside surface of the micelle, which is surrounded by water, and the hydrocarbon tails are sequestered inside, interacting with one another (Figure 12.9).

Alternatively, the strongly opposed preferences of the hydrophilic and hydrophobic moieties of membrane lipids can be satisfied by forming a *lipid bilayer,* composed of two lipid sheets (Figure 12.10). A lipid bilayer is also called a *bimolecular sheet.* The hydrophobic tails of each individual sheet interact with one another, forming a hydrophobic interior that acts as a permeability barrier. The hydrophilic head groups interact with the aqueous medium on each side of the bilayer. The two opposing sheets are called leaflets.

The favored structure for most phospholipids and glycolipids in aqueous media is a bimolecular sheet rather than a micelle. The reason is that the two fatty acid chains of a phospholipid or a glycolipid are too bulky to fit into the interior of a micelle. In contrast, salts of fatty acids (such as sodium palmitate, a constituent of soap) readily form micelles because they contain only one chain. *The formation of bilayers instead of micelles by phospholipids is of critical biological importance.* A micelle is a limited structure, usually less than 200 Å (20 nm) in diameter. In contrast, a bimolecular sheet can extend to macroscopic dimensions, as much as a millimeter (10^7 Å, or 10^6 nm) or more. Phospholipids and related molecules are important membrane constituents because they readily form extensive bimolecular sheets.

Lipid bilayers form spontaneously by a *self-assembly process.* In other words, the structure of a bimolecular sheet is inherent in the structure of the constituent lipid molecules. The growth of lipid bilayers from phospholipids is rapid and spontaneous in water. *Hydrophobic interactions are the major driving force for the formation of lipid bilayers.* Recall that hydrophobic interactions also play a dominant role in the stacking of bases in nucleic acids and in the folding of proteins (Sections 1.3 and 2.4). Water molecules are released from the hydrocarbon tails of membrane lipids as these tails become sequestered in the nonpolar interior of the bilayer. Furthermore, *van der Waals attractive forces between the hydrocarbon tails favor close packing of the tails*. Finally, there are *electrostatic and hydrogen-bonding attractions between the polar head groups and water molecules.* Thus, lipid bilayers are stabilized by the full array of forces that mediate molecular interactions in biological systems. Because lipid bilayers are held together by many *reinforcing, noncovalent interactions (predominantly hydrophobic)*, they are *cooperative structures.* These hydrophobic interactions have three significant biological consequences: (1) lipid bilayers have an inherent tendency to be *extensive;* (2) lipid bilayers will tend to *close on themselves* so that there are no edges with exposed hydrocarbon chains, and so they form compartments; and (3) lipid bilayers are *self-sealing* because a hole in a bilayer is energetically unfavorable.

Lipid vesicles can be formed from phospholipids

The propensity of phospholipids to form membranes has been used to create an important experimental and clinical tool. *Lipid vesicles,* or *liposomes,* are aqueous compartments enclosed by a lipid bilayer (Figure 12.11). These structures can be used to study membrane permeability or to deliver chemicals

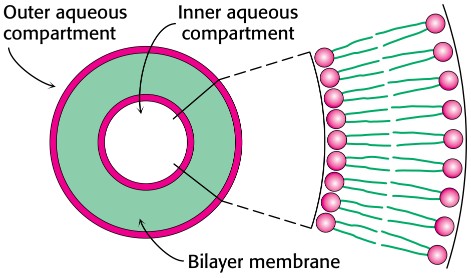


FIGURE 12.11 Liposome. A liposome, or lipid vesicle, is a small aqueous compartment surrounded by a lipid bilayer.

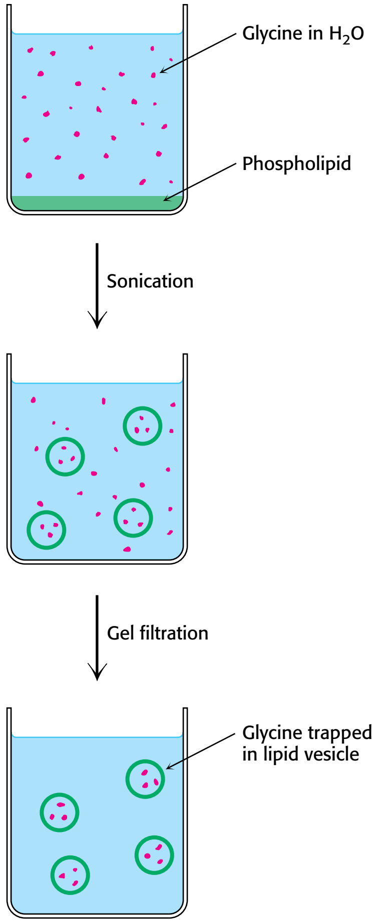


FIGURE 12.12 Preparation of glycine-containing liposomes. Liposomes containing glycine are formed by the sonication of phospholipids in the presence of glycine. Free glycine is removed by gel filtration.

to cells. Liposomes are formed by suspending a suitable lipid, such as phosphatidylcholine, in an aqueous medium, and then *sonicating* (i.e., agitating by high-frequency sound waves) to give a dispersion of closed vesicles that are quite uniform in size. Vesicles formed by this method are nearly spherical and have a diameter of about 500 Å (50 nm). Larger vesicles (of the order of 1 μm or 10^4 Å in diameter) can be prepared by slowly evaporating the organic solvent from a suspension of phospholipid in a mixed-solvent system.

Ions or molecules can be trapped in the aqueous compartments of lipid vesicles by forming the vesicles in the presence of these substances (Figure 12.12). For example, 500-Å-diameter vesicles formed in a 0.1 M glycine solution will trap about 2000 molecules of glycine in each inner aqueous compartment. These glycine-containing vesicles can be separated from the surrounding solution of glycine by dialysis or by gel-filtration chromatography (Section 3.1). The permeability of the bilayer membrane to glycine can then be determined by measuring the rate of efflux of glycine from the inner compartment of the vesicle to the ambient solution. Liposomes can be formed with specific membrane proteins embedded in them by solubilizing the proteins in the presence of detergents and then adding them to the phospholipids from which liposomes will be formed. Protein–liposome complexes provide valuable experimental tools for examining a range of membrane-protein functions.

Therapeutic applications for liposomes are currently under active investigation. For example, liposomes containing drugs or DNA can be injected into patients. These liposomes fuse with the plasma membrane of many kinds of cells, introducing into the cells the molecules that they contain. Drug delivery with liposomes often lessens its toxicity. Less of the drug is distributed to normal tissues because long-circulating liposomes concentrate in regions of increased blood circulation, such as solid tumors and sites of inflammation. Moreover, the selective fusion of lipid vesicles with particular kinds of cells is a promising means of controlling the delivery of drugs to target cells.

Another well-defined synthetic membrane is a *planar bilayer membrane.* This structure can be formed across a 1-mm hole in a partition between two aqueous compartments by dipping a fine paintbrush into a membrane-forming solution, such as phosphatidylcholine in decane, and stroking the tip of the brush across the hole. The lipid film across the hole thins spontaneously into a lipid bilayer. The electrical conduction properties of this macroscopic bilayer membrane are readily studied by inserting electrodes into each aqueous compartment (Figure 12.13). For example, the permeability of the membrane to ions is determined by measuring the current across the membrane as a function of the applied voltage.

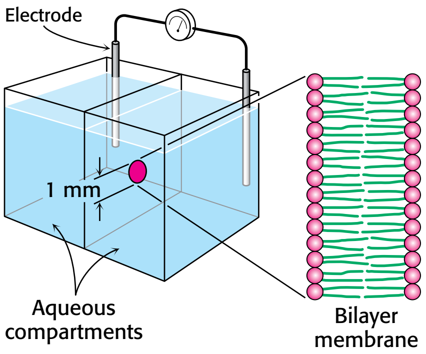


FIGURE 12.13 Experimental arrangement for the study of a planar bilayer membrane. A bilayer membrane is formed across a 1-mm hole in a septum that separates two aqueous compartments. This arrangement permits measurements of the permeability and electrical conductance of lipid bilayers.

Lipid bilayers are highly impermeable to ions and most polar molecules

Permeability studies of lipid vesicles and electrical-conductance measurements of planar bilayers have shown that *lipid bilayer membranes have a very low permeability for ions and most polar molecules.* Water is a conspicuous

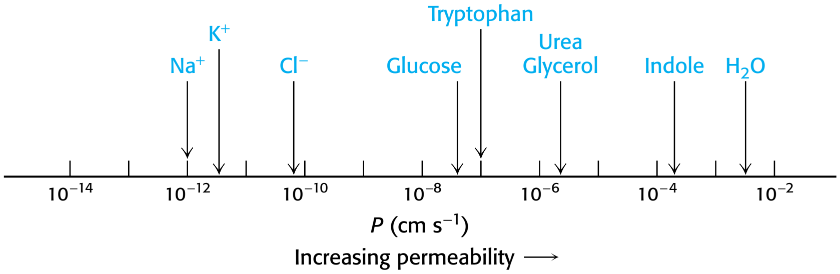


FIGURE 12.14 Permeability coefficients (*P*) of ions and molecules in a lipid bilayer. The ability of molecules to cross a lipid bilayer spans a wide range of values.

exception to this generalization; it traverses such membranes relatively easily because of its low molecular weight, high concentration, and lack of a complete charge. The range of measured permeability coefficients is very wide (Figure 12.14). For example, Na^+ and K^+ traverse these membranes 10^9 times as slowly as does H_2O. Tryptophan, a zwitterion at pH 7, crosses the membrane 10^3 times as slowly as does indole, a structurally related molecule that lacks ionic groups. In fact, *the permeability of small molecules is correlated with their solubility in a nonpolar solvent relative to their solubility in water.* This relation suggests that a small molecule might traverse a lipid bilayer membrane in the following way: first, it sheds its solvation shell of water; then, it is dissolved in the hydrocarbon core of the membrane; and, finally, it diffuses through this core to the other side of the membrane, where it becomes resolvated by water. An ion such as Na^+ traverses membranes very slowly because the replacement of its coordination shell of polar water molecules by nonpolar interactions with the membrane interior is highly unfavorable energetically.

12.4 Proteins Carry Out Most Membrane Processes

We now turn to membrane proteins, which are responsible for most of the dynamic processes carried out by membranes. Membrane lipids form a permeability barrier and thereby establish compartments, whereas *specific proteins mediate nearly all other membrane functions.* In particular, proteins transport chemicals and information across a membrane. Membrane lipids create the appropriate environment for the action of such proteins.

Membranes differ in their protein content. Myelin, a membrane that serves as an electrical insulator around certain nerve fibers, has a low content of protein (18%). Relatively pure lipids are well suited for insulation. In contrast, the plasma membranes, or exterior membranes, of most other cells are much more metabolically active. They contain many pumps, channels, receptors, and enzymes. The protein content of these plasma membranes is typically 50%. Energy-transduction membranes, such as the internal membranes of mitochondria and chloroplasts, have the highest content of protein, around 75%.

The protein components of a membrane can be readily visualized by *SDS–polyacrylamide gel electrophoresis.* As stated earlier (p. 73), the electrophoretic mobility of many proteins in SDS-containing gels depends on the mass rather than on the net charge of the protein. The gel-electrophoresis patterns of three membranes—the plasma membrane of erythrocytes, the photoreceptor membrane of retinal rod cells, and the sarcoplasmic reticulum membrane of muscle—are shown in Figure 12.15. It is evident that each of these three membranes contains many proteins but has a distinct protein composition. In general, *membranes performing different functions contain different repertoires of proteins.*

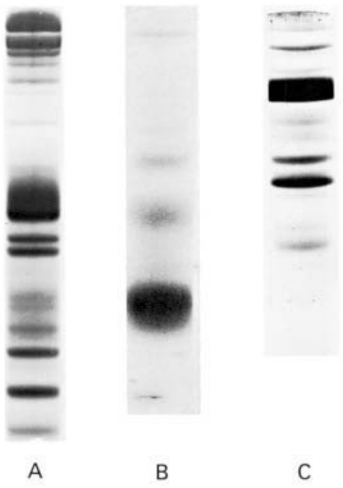


FIGURE 12.15 SDS–acrylamide gel patterns of membrane proteins. (A) The plasma membrane of erythrocytes. (B) The photoreceptor membranes of retinal rod cells. (C) The sarcoplasmic reticulum membrane of muscle cells. [Courtesy of Dr. Theodore Steck (part A) and Dr. David MacLennan (part C).]

Proteins associate with the lipid bilayer in a variety of ways

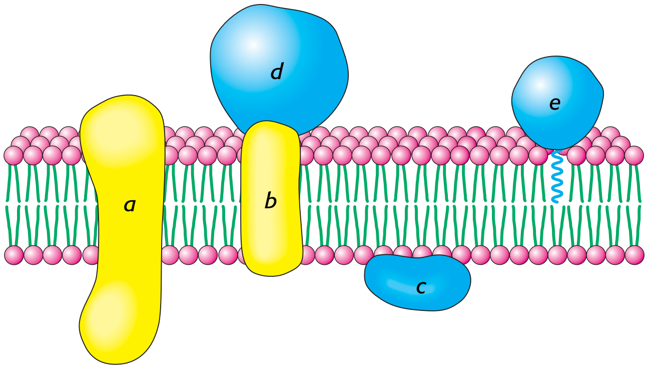


FIGURE 12.16 Integral and peripheral membrane proteins. Integral membrane proteins (*a* and *b*) interact extensively with the hydrocarbon region of the bilayer. Most known integral membrane proteins traverse the lipid bilayer. Peripheral membrane proteins interact with the polar head groups of the lipids (*c*) or bind to the surfaces of integral proteins (*d*). Other proteins are tightly anchored to the membrane by a covalently attached lipid molecule (*e*).

The ease with which a protein can be dissociated from a membrane indicates how intimately it is associated with the membrane. Some membrane proteins can be solubilized by relatively mild means, such as extraction by a solution of high ionic strength (e.g., 1 M NaCl). Other membrane proteins are bound much more tenaciously; they can be solubilized only by using a detergent or an organic solvent. Membrane proteins can be classified as being either *peripheral* or *integral* on the basis of this difference in dissociability (Figure 12.16). *Integral membrane proteins* interact extensively with the hydrocarbon chains of membrane lipids, and they can be released only by agents that compete for these nonpolar interactions. In fact, most integral membrane proteins span the lipid bilayer. In contrast, *peripheral membrane proteins* are bound to membranes primarily by electrostatic and hydrogen-bond interactions with the head groups of lipids. These polar interactions can be disrupted by adding salts or by changing the pH. Many peripheral membrane proteins are bound to the surfaces of integral proteins, on either the cytoplasmic or the extracellular side of the membrane. Others are anchored to the lipid bilayer by a covalently attached hydrophobic chain, such as a fatty acid.

Proteins interact with membranes in a variety of ways

Membrane proteins are more difficult to purify and crystallize than are water-soluble proteins. Nonetheless, researchers using x-ray crystallographic or electron microscopic methods have determined the three-dimensional structures of more than 2000 such proteins at sufficiently high resolution to discern the molecular details. As noted in Chapter 2, membrane proteins differ from soluble proteins in the distribution of hydrophobic and hydrophilic groups. We will consider the structures of three membrane proteins in some detail.

Proteins can span the membrane with alpha helices. The first membrane protein that we consider is the archaeal protein *bacteriorhodopsin*, shown in Figure 12.17. This protein uses light energy to transport protons from inside to outside the cell, generating a proton gradient used to form ATP.

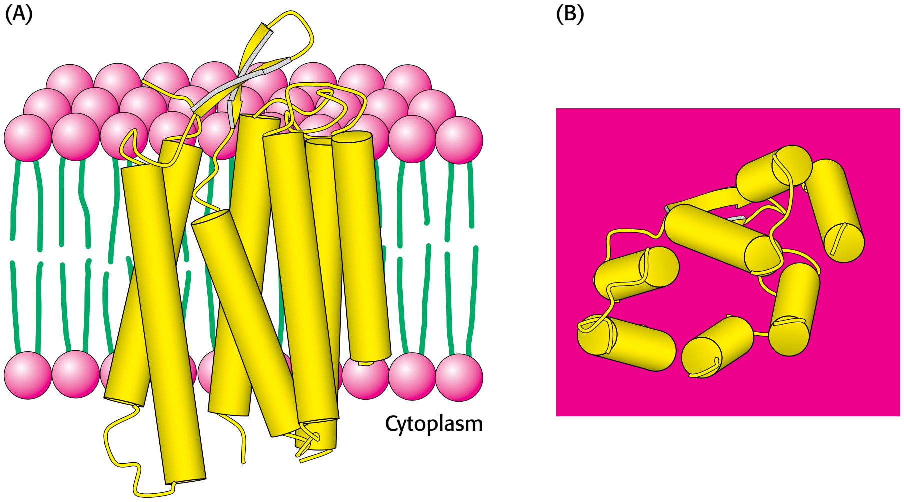


FIGURE 12.17 Structure of bacteriorhodopsin. *Notice* that bacteriorhodopsin consists largely of membrane-spanning α helices (represented by yellow cylinders). (A) View through the membrane bilayer. The interior of the membrane is green and the head groups are red. (B) View from the cytoplasmic side of the membrane. [Drawn from 1BRX.pdb.]

A Q I T G R P E W I W L A L G T A L M G L G T L Y F L V K G M G V S D P D A K K F Y A I T T L V P A
I A F T M Y L S M L L G Y G L T M V P F G G E Q N P I Y W A R Y A D W L F T T P L L L L D L A L L V
D A D Q G T I L A L V G A D G I M I G T G L V G A L T K V Y S Y R F V W W A I S T A A M L Y I L Y V
L F F G F T S K A E S M R P E V A S T F K V L R N V T V V L W S A Y V V V W L I G S E G A G I V P L
N I E T L L F M V L D V S A K V G F G L I L L R S R A I F G E A E A P E P S A D G A A A T S

FIGURE 12.18 Amino acid sequence of bacteriorhodopsin. The seven helical regions are highlighted in yellow and the charged residues in red.

Bacteriorhodopsin is built almost entirely of α helices; seven closely packed α helices, nearly perpendicular to the plane of the cell membrane, span its 45-Å width. Examination of the primary structure of bacteriorhodopsin reveals that most of the amino acids in these membrane-spanning α helices are nonpolar and only a very few are charged (Figure 12.18). This distribution of nonpolar amino acids is sensible because these residues are either in contact with the hydrocarbon core of the membrane or with one another. *Membrane-spanning α helices are the most common structural motif in membrane proteins.* As will be considered in Section 12.5, such regions can often be detected by examining amino acid sequence alone.

A channel protein can be formed from beta strands. Porin, a protein from the outer membranes of bacteria such as *E. coli* and *Rhodobacter capsulatus,* represents a class of membrane proteins with a completely different type of structure. Structures of this type are built from β strands and contain essentially no α helices (Figure 12.19).

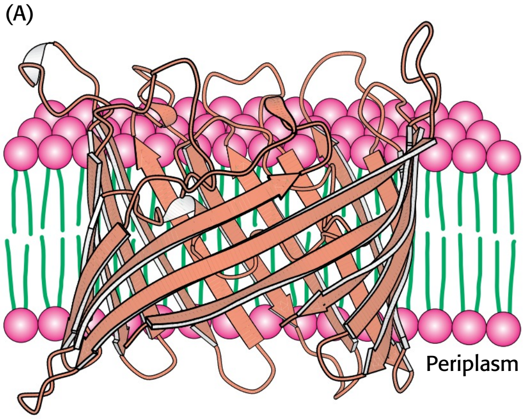


FIGURE 12.19 Structure of bacterial porin (from *Rhodopseudomonas blastica*). *Notice* that this membrane protein is built entirely of β strands. (A) Side view. (B) View from the periplasmic space. Only one monomer of the trimeric protein is shown. [Drawn from 1PRN.pdb.]

The arrangement of β strands is quite simple: each strand is hydrogen bonded to its neighbor in an antiparallel arrangement, forming a single β sheet. The β sheet curls up to form a hollow cylinder that, as its name suggests, forms a pore, or channel, in the membrane. The outside surface of porin is appropriately nonpolar, given that it interacts with the hydrocarbon core of the membrane. In contrast, the inside of the channel is quite hydrophilic and is filled with water. This arrangement of nonpolar and polar surfaces is accomplished by the alternation of hydrophobic and hydrophilic amino acids along each β strand (Figure 12.20).

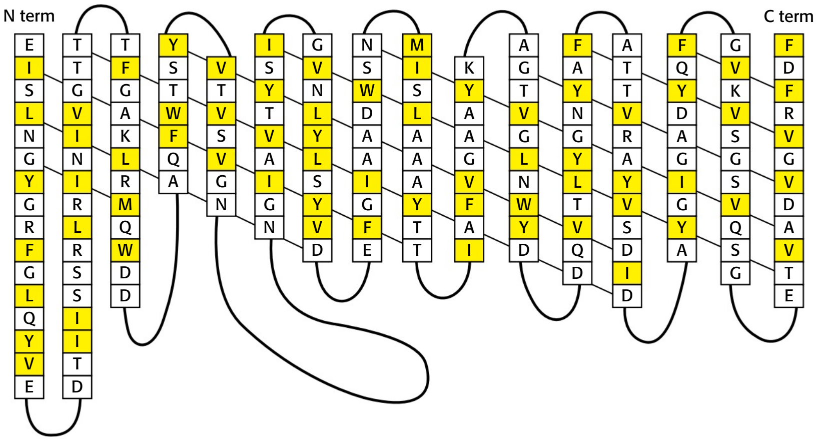


FIGURE 12.20 Amino acid sequence of a porin. Some membrane proteins, such as porins, are built from β strands that tend to have hydrophobic and hydrophilic amino acids in adjacent positions. The secondary structure of porin from *Rhodopseudomonas blastica* is shown, with the diagonal lines indicating the direction of hydrogen bonding along the β sheet. Hydrophobic residues (F, I, L, M, V, W, and Y) are shown in yellow. These residues tend to lie on the outside of the structure, in contact with the hydrophobic core of the membrane.

Embedding part of a protein in a membrane can link the protein to the membrane surface. The structure of the endoplasmic reticulum membrane-bound enzyme prostaglandin H_2 synthase-1 reveals a rather different role for α helices in protein–membrane associations. This enzyme catalyzes the conversion of arachidonic acid into prostaglandin H_2 in two steps: (1) a cyclooxygenase reaction and (2) a peroxidase reaction (Figure 12.21). Prostaglandin H_2 promotes inflammation and modulates gastric acid secretion. The enzyme that produces prostaglandin H_2 is a homodimer with a rather complicated structure consisting primarily of α helices. Unlike bacteriorhodopsin, this protein is not largely embedded in the membrane. Instead, it lies along the outer surface of the membrane, firmly bound by a set of α helices with hydrophobic surfaces that extend from the bottom of the protein into the membrane (Figure 12.22). This linkage is sufficiently strong that only the action of detergents can release the protein from the membrane. Thus, this enzyme is classified as an integral membrane protein, although it does not span the membrane.

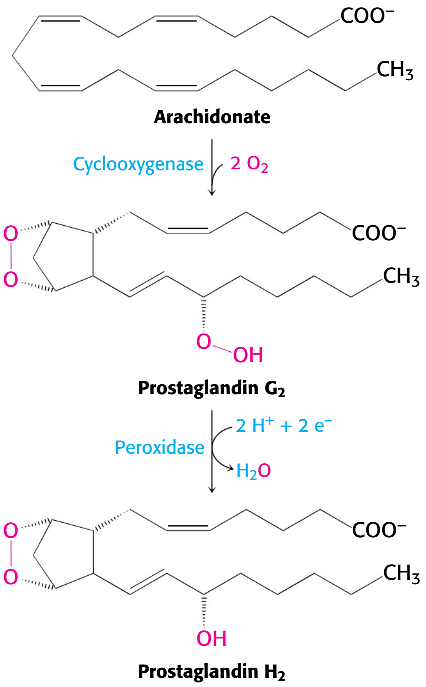


FIGURE 12.21 Formation of prostaglandin H_2. Prostaglandin H_2 synthase-1 catalyzes the formation of prostaglandin H_2 from arachidonic acid in two steps.

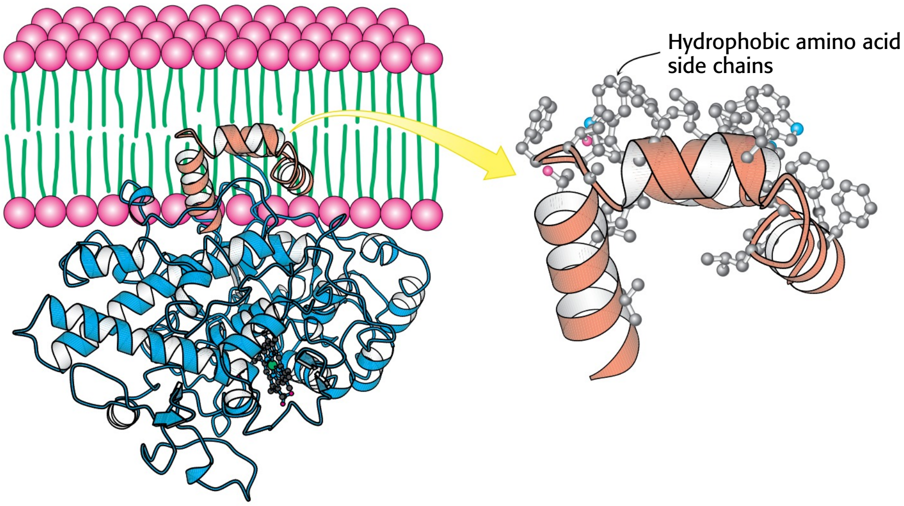


FIGURE 12.22 Attachment of prostaglandin H_2 synthase-1 to the membrane. *Notice* that prostaglandin H_2 synthase-1 is held in the membrane by a set of α helices (orange) coated with hydrophobic side chains. One monomer of the dimeric enzyme is shown. [Drawn from 1PTH.pdb.]

The localization of prostaglandin H_2 synthase-l in the membrane is crucial to its function. The substrate for this enzyme, arachidonic acid, is a hydrophobic molecule generated by the hydrolysis of membrane lipids. Arachidonic acid reaches the active site of the enzyme from the membrane without entering an aqueous environment by traveling through a hydrophobic channel in the protein (Figure 12.23). Indeed, nearly all of us have experienced the importance of this channel: drugs such as aspirin and ibuprofen block the channel and prevent prostaglandin synthesis by inhibiting the cyclooxygenase activity of the synthase. In particular, aspirin acts through the transfer of its acetyl group to a serine residue (Ser 530) that lies along the path to the active site (Figure 12.24).

Two important features emerge from our examination of these three examples of membrane-protein structure.

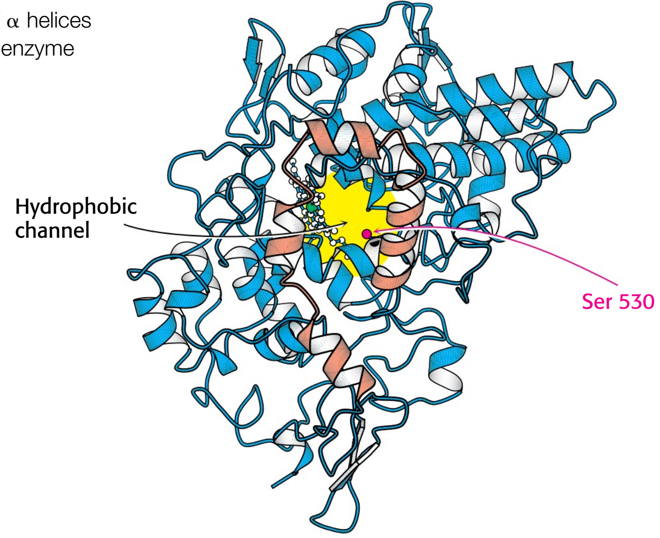
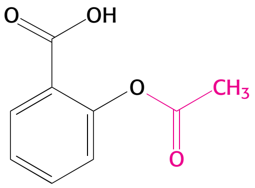


FIGURE 12.23 Hydrophobic channel of prostaglandin H_2 synthase-1. A view of prostaglandin H_2 synthase-1 from the membrane shows the hydrophobic channel that leads to the active site. The membrane-anchoring helices are shown in orange. [Drawn from 1PTH.pdb.]



Aspirin
(Acetylsalicyclic acid)

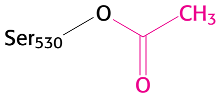
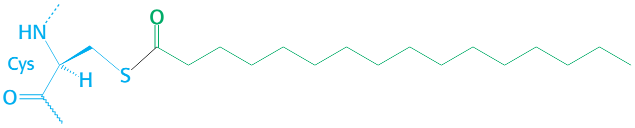


FIGURE 12.24 Aspirin's effects on prostaglandin H_2 synthase-1. Aspirin acts by transferring an acetyl group to a serine residue in prostaglandin H_2 synthase-1.

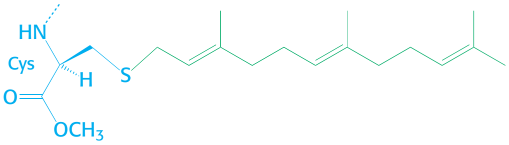
First, the parts of the protein that interact with the hydrophobic parts of the membrane are coated with nonpolar amino acid side chains, whereas those parts that interact with the aqueous environment are much more hydrophilic. Second, the structures positioned within the membrane are quite regular and, in particular, all backbone hydrogen-bond donors and acceptors participate in hydrogen bonds. *Breaking a hydrogen bond within a membrane is quite unfavorable, because little or no water is present to compete for the polar groups.*

Some proteins associate with membranes through covalently attached hydrophobic groups

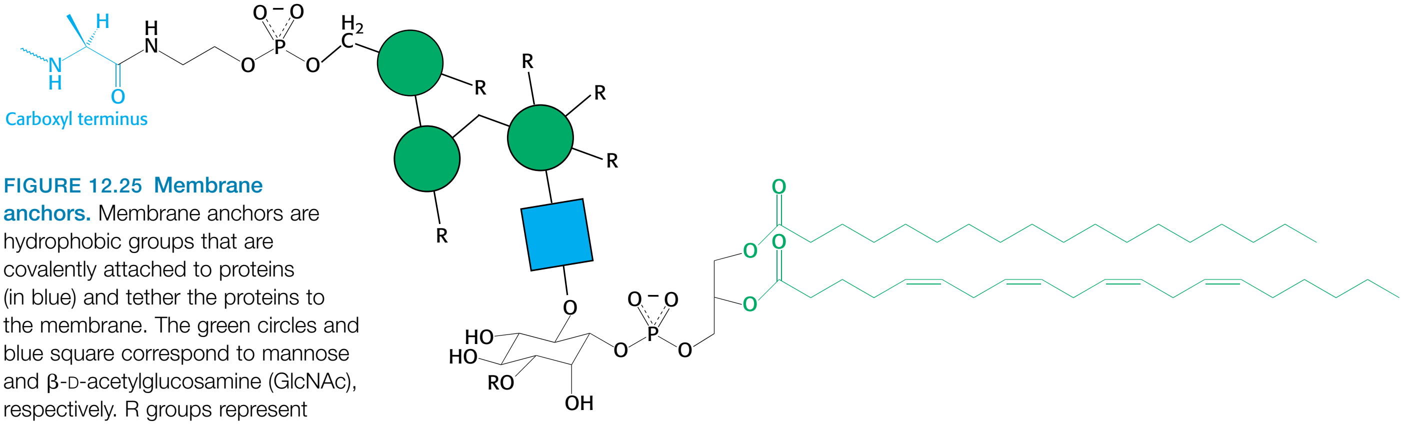
The membrane proteins considered thus far associate with the membrane through surfaces generated by hydrophobic amino acid side chains. However, even otherwise soluble proteins can associate with membranes if hydrophobic groups are attached to the proteins. Three such groups are shown in Figure 12.25: (1) a palmitoyl group attached to a specific cysteine residue by a thioester bond, (2) a farnesyl group attached to a cysteine residue at the carboxyl terminus, and (3) a glycolipid structure termed a glycosylphosphatidylinositol (GPI) anchor attached to the carboxyl terminus. These modifications are attached by enzyme systems that recognize specific signal sequences near the site of attachment.



_S_-Palmitoylcysteine



C-terminal _S_-farnesylcysteine methyl ester



Glycosylphosphatidylinositol (GPI) anchor

FIGURE 12.25 Membrane anchors. Membrane anchors are hydrophobic groups that are covalently attached to proteins (in blue) and tether the proteins to the membrane. The green circles and blue square correspond to mannose and β-D-acetylglucosamine (GlcNAc), respectively. R groups represent points of additional modification.

Transmembrane helices can be accurately predicted from amino acid sequences

Many membrane proteins, like bacteriorhodopsin, employ α helices to span the hydrophobic part of a membrane. As noted earlier, most of the residues in these α helices are nonpolar and almost none of them are charged. Can we use this information to identify likely membrane-spanning regions from sequence data alone? One approach to identifying transmembrane helices is to ask whether a postulated helical segment is likely to be more stable in a hydrocarbon environment or in water. Specifically, we want to estimate the free-energy change when a helical segment is transferred from the interior of a membrane to water. Free-energy changes for the transfer of individual amino acid residues from a hydrophobic to an aqueous environment are given in Table 12.2. For example, the transfer of a helix formed entirely of L-arginine residues, a positively charged amino acid, from the interior of a membrane to water would be highly favorable [−51.5 kJ mol^{-1} (−12.3 kcal mol^{-1}) per arginine residue in the helix]. In contrast, the transfer of a helix formed entirely of L-phenylalanine, a hydrophobic amino acid, would be unfavorable [+15.5 kJ mol^{-1} (+3.7 kcal mol^{-1}) per phenylalanine residue in the helix].

The hydrocarbon core of a membrane is typically 30 Å wide, a length that can be traversed by an α helix consisting of 20 residues. We can take the amino acid sequence of a protein and estimate the free-energy change that takes place when a hypothetical α helix formed of residues 1 through 20 is transferred from the membrane interior to water. The same calculation can be made for residues 2 through 21, 3 through 22, and so forth, until we reach the end of the sequence. The span of 20 residues chosen for this calculation is called a *window*. The free-energy change for each window is plotted against the first amino acid at the window to create a *hydropathy plot*. Empirically, a peak of +84 kJ mol^{-1} (+20 kcal mol^{-1}) or more in a hydropathy plot based on a window of 20 residues indicates that a polypeptide segment could be a membrane-spanning α helix. For example, glycophorin, a protein found in the membranes of red blood cells, is predicted by this criterion to have one membrane-spanning helix, in agreement with experimental findings (Figure 12.26). Note, however, that a peak in the hydropathy plot does not prove that a segment is a transmembrane helix. Even soluble proteins may have highly nonpolar regions. Conversely, some membrane proteins contain membrane-spanning features (such as

TABLE 12.2 Polarity scale for identifying transmembrane helices

Amino acid residue	Transfer free energy in kJ mol^{-1} (kcal mol^{-1})
Phe	15.5 (3.7)
Met	14.3 (3.4)
Ile	13.0 (3.1)
Leu	11.8 (2.8)
Val	10.9 (2.6)
Cys	8.4 (2.0)
Trp	8.0 (1.9)
Ala	6.7 (1.6)
Thr	5.0 (1.2)
Gly	4.2 (1.0)
Ser	2.5 (0.6)
Pro	−0.8 (−0.2)
Tyr	−2.9 (−0.7)
His	−12.6 (−3.0)
Gln	−17.2 (−4.1)
Asn	−20.2 (−4.8)
Glu	−34.4 (−8.2)
Lys	−37.0 (−8.8)
Asp	−38.6 (−9.2)
Arg	−51.7 (−12.3)

Source: Data from D. M. Engelman, T. A. Steitz, and A. Goldman. Annu. Rev. Biophys. Biophys. Chem. 15(1986):321–353.

Note: The free energies are for the transfer of an amino acid residue in an α helix from the membrane interior (assumed to have a dielectric constant of 2) to water.

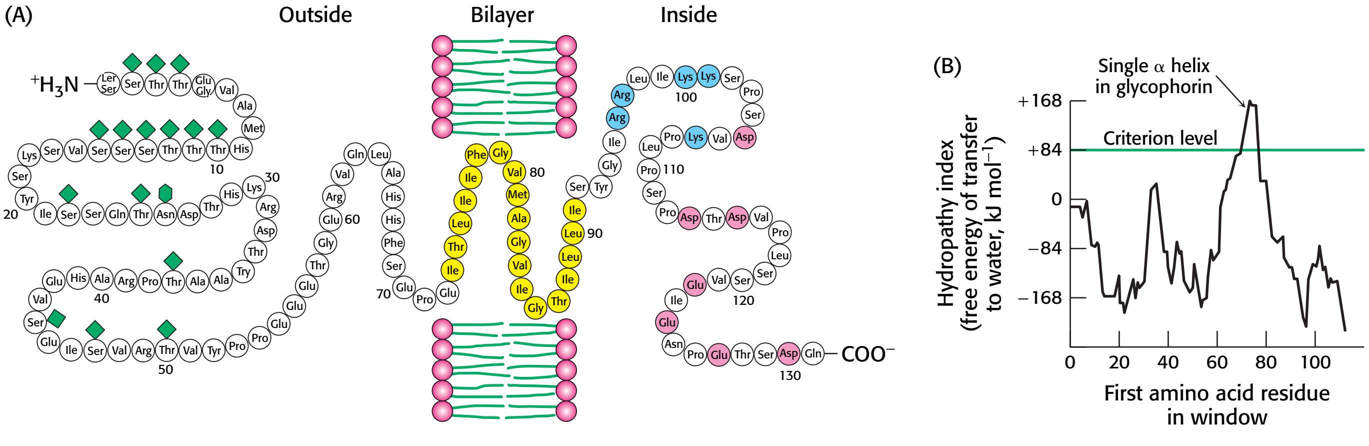
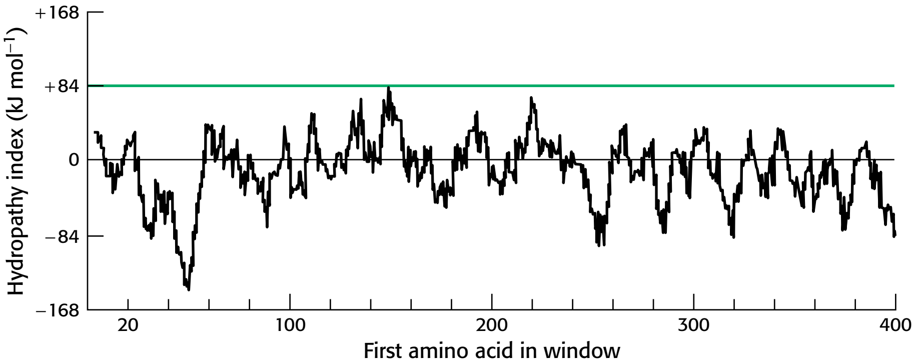

(A)
Outside
Bilayer
Inside
$^{+}H_3N$
COO^{-}
10
20
30
40
50
60
70
80
90
100
110
120
130
(B)
Single α helix in glycophorin
Criterion level
Hydropathy index (free energy of transfer to water, kJ mol^{-1})
+168
+84
0
−84
−168
0
20
40
60
80
100
First amino acid residue in window

FIGURE 12.26 Locating the membrane-spanning helix of glycophorin. (A) Amino acid sequence and transmembrane disposition of glycophorin A from the red-blood-cell membrane. Fifteen *O*-linked carbohydrate units are shown as diamond shapes, and an *N*-linked unit is shown as a lozenge shape. The hydrophobic residues (yellow) buried in the bilayer form a transmembrane α helix. The carboxyl-terminal part of the molecule, located on the cytoplasmic side of the membrane, is rich in negatively charged (red) and positively charged (blue) residues. (B) Hydropathy plot for glycophorin. The free energy for transferring a helix of 20 residues from the membrane to water is plotted as a function of the position of the first residue of the helix in the sequence of the protein. Peaks of greater than +84 kJ mol^{-1} (+20 kcal mol^{-1}) in hydropathy plots are indicative of potential transmembrane helices. [(A) Information from Dr. Vincent Marchesi; (B) data from D. M. Engelman, T. A. Steitz, and A. Goldman, *Annu. Rev. Biophys. Biophys. Chem*. 15:321–353, 1986. Copyright © 1986 by Annual Reviews, Inc. All rights reserved.]

FIGURE 12.27 Hydropathy plot for porin. No strong peaks are observed for this intrinsic membrane protein, because it is constructed from membrane-spanning β strands rather than α helices.



a set of cylinder-forming β strands) that escape detection by these plots (Figure 12.27).

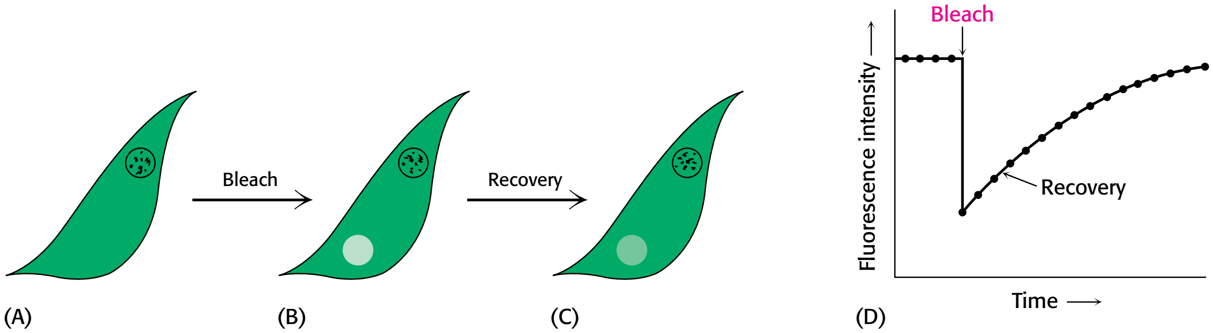
12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane

Biological membranes are not rigid, static structures. On the contrary, lipids and many membrane proteins are constantly in lateral motion, a process called *lateral diffusion*. The rapid lateral movement of membrane proteins has been visualized by means of fluorescence microscopy using the technique of *fluorescence recovery after photobleaching* (FRAP; Figure 12.28). First, a cell-surface component is specifically labeled with a fluorescent chromophore. A small region of the cell surface (~3 μm^2) is viewed through a fluorescence microscope. The fluorescent molecules in this region are then destroyed (bleached) by a very intense light pulse from a laser, as indicated by the pale spot in Figure 12.28B. The fluorescence of this region is subsequently monitored as a function of time by using a light level sufficiently low to prevent further bleaching. If the labeled component is mobile, bleached molecules leave and unbleached molecules enter the illuminated region, resulting in an increase in the fluorescence intensity. The rate of recovery of fluorescence depends on the lateral mobility of the fluorescence-labeled component, which can be expressed in terms of a diffusion coefficient, D. The average distance S traversed in time t depends on D according to the expression

$$S = (4Dt)^{1/2}$$

The diffusion coefficient of lipids in a variety of membranes is about 1 $\mu m^2\ s^{-1}$. Thus, a phospholipid molecule diffuses an average distance of 2 μm in 1 s. This rate means that *a lipid molecule can travel from one end of a bacterium to the other in a second*. The magnitude of the observed diffusion coefficient indicates that the viscosity of the membrane is about 100 times that of water, rather like that of olive oil.

FIGURE 12.28 Fluorescence recovery after photobleaching (FRAP) technique. (A) The cell surface fluoresces because of a labeled surface component. (B) The fluorescent molecules of a small part of the surface are bleached by an intense light pulse. (C) The fluorescence intensity recovers as bleached molecules diffuse out of the region and unbleached molecules diffuse into it. (D) The rate of recovery depends on the diffusion coefficient.



In contrast, proteins vary markedly in their lateral mobility. *Some proteins are nearly as mobile as lipids, whereas others are virtually immobile.* For example, the photoreceptor protein rhodopsin (Section 33.3), a very mobile protein, has a diffusion coefficient of 0.4 $\mu m^2\ s^{-1}$. The rapid movement of rhodopsin is essential for fast signaling. At the other extreme is fibronectin, a peripheral glycoprotein that interacts with the extracellular matrix. For fibronectin, D is less than $10^{-4}\ \mu m^2\ s^{-1}$. Fibronectin has a very low mobility because it is anchored to actin filaments on the inside of the plasma membrane through *integrin,* a transmembrane protein that links the extracellular matrix to the cytoskeleton.

The fluid mosaic model allows lateral movement but not rotation through the membrane

On the basis of the mobility of proteins in membranes, in 1972 S. Jonathan Singer and Garth Nicolson proposed a *fluid mosaic model* to describe the overall organization of biological membranes. The essence of their model is that *membranes are two-dimensional solutions of oriented lipids and globular proteins.* The lipid bilayer has a dual role: it is both a *solvent* for integral membrane proteins and a *permeability barrier.* Membrane proteins are free to diffuse laterally in the lipid matrix unless restricted by special interactions.

Although the lateral diffusion of membrane components can be rapid, the spontaneous rotation of lipids from one face of a membrane to the other is a very slow process. The transition of a molecule from one membrane surface to the other is called *transverse diffusion* or *flip-flop* (Figure 12.29). The flip-flop of phospholipid molecules in phosphatidylcholine vesicles has been directly measured by electron spin resonance techniques, which show that *a phospholipid molecule flip-flops once in several hours.* Thus, a phospholipid molecule takes about 10^9 times as long to flip-flop across a membrane as it takes to diffuse a distance of 50 Å in the lateral direction. The free-energy barriers to flip-flopping are even larger for protein molecules than for lipids because proteins have more-extensive polar regions. In fact, the flip-flop of a protein molecule has not been observed. Hence, *membrane asymmetry can be preserved for long periods.*

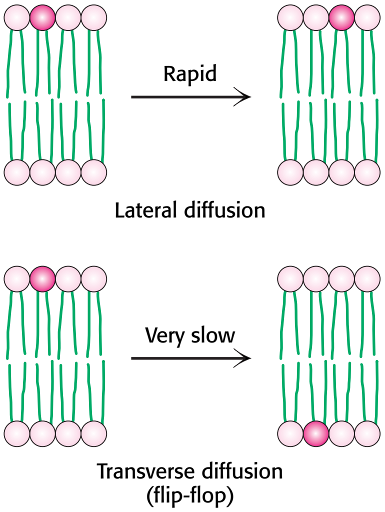


FIGURE 12.29 Lipid movement in membranes. Lateral diffusion of lipids is much more rapid than transverse diffusion (flip-flop).

Membrane fluidity is controlled by fatty acid composition and cholesterol content

Many membrane processes, such as transport or signal transduction, depend on the fluidity of the membrane lipids, which in turn depends on the properties of fatty acid chains. Fatty acid chains in membrane bilayers can exist in an ordered, rigid state or in a relatively disordered, fluid state. The transition from the rigid to the fluid state takes place abruptly as the temperature is raised above T_m, the melting temperature (Figure 12.30). *This transition temperature depends on the length of the fatty acid chains and on their degree of unsaturation* (Table 12.3). The presence of saturated fatty

TABLE 12.3 The melting temperature of phosphatidylcholine containing different pairs of identical fatty acid chains

		Fatty acid		
Number of carbons	Number of double bonds	Common name	Systematic name	T_m (°C)
22	0	Behenate	*n*-Docosanote	75
18	0	Stearate	*n*-Octadecanoate	58
16	0	Palmitate	*n*-Hexadecanoate	41
14	0	Myristate	*n*-Tetradecanoate	24
18	1	Oleate	*cis*-Δ^9-Octadecenoate	−22

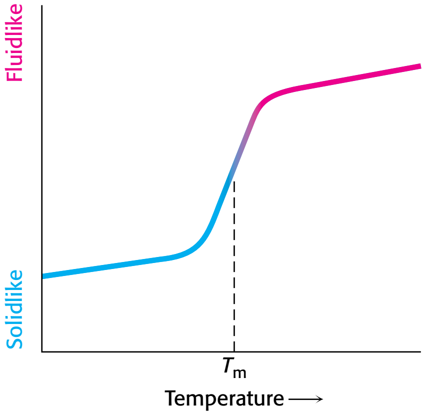


FIGURE 12.30 The phase-transition, or melting, temperature (T_m) for a phospholipid membrane. As the temperature is raised, the phospholipid membrane changes from a packed, ordered state to a more random one.

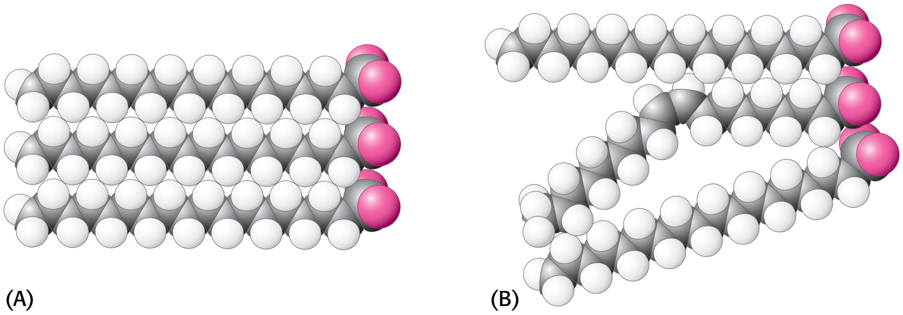


FIGURE 12.31 Packing of fatty acid chains in a membrane. The highly ordered packing of fatty acid chains is disrupted by the presence of cis double bonds. The space-filling models show the packing of (A) three molecules of stearate (C_{18}, saturated) and (B) a molecule of oleate (C_{18}, unsaturated) between two molecules of stearate.

acid residues favors the rigid state because their straight hydrocarbon chains interact very favorably with one another. On the other hand, *a cis double bond produces a bend in the hydrocarbon chain. This bend interferes with a highly ordered packing of fatty acid chains, and so* T_m *is lowered* (Figure 12.31). The length of the fatty acid chain also affects the transition temperature. Long hydrocarbon chains interact more strongly than do short ones. Specifically, each additional —CH_2— group makes a favorable contribution of about -2 kJ mol^{-1} (-0.5 kcal mol^{-1}) to the free energy of interaction of two adjacent hydrocarbon chains.

Bacteria regulate the fluidity of their membranes by varying the number of double bonds and the length of their fatty acid chains. For example, the ratio of saturated to unsaturated fatty acid chains in the *E. coli* membrane decreases from 1.6 to 1.0 as the growth temperature is lowered from 42°C to 27°C. This decrease in the proportion of saturated residues prevents the membrane from becoming too rigid at the lower temperature.

In animals, cholesterol is the key regulator of membrane fluidity. Cholesterol contains a bulky steroid nucleus with a hydroxyl group at one end and a flexible hydrocarbon tail at the other end. Cholesterol inserts into bilayers with its long axis perpendicular to the plane of the membrane. The hydroxyl group of cholesterol forms a hydrogen bond with a carbonyl oxygen atom of a phospholipid head group, whereas the hydrocarbon tail of cholesterol is located in the nonpolar core of the bilayer. The different shape of cholesterol compared with that of phospholipids disrupts the regular interactions between fatty acid chains (Figure 12.32).

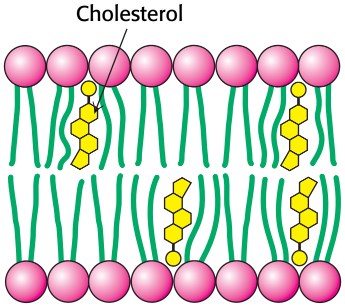


FIGURE 12.32 Cholesterol disrupts the tight packing of the fatty acid chains. [Information from on S. L. Wolfe, *Molecular and Cellular Biology* (Wadsworth, 1993).]

Lipid rafts are highly dynamic complexes formed between cholesterol and specific lipids

In addition to its nonspecific effects on membrane fluidity, cholesterol can form specific complexes with lipids that contain the sphingosine backbone, including sphingomyelin and certain glycolipids, and with GPI-anchored proteins. These complexes concentrate within small (10–200 nm) and highly dynamic regions within membranes. The resulting structures are often referred to as *lipid rafts*. One result of these interactions is the *moderation of membrane fluidity,* making membranes less fluid but at the same time less subject to phase transitions. The presence of lipid rafts thus represents a modification of the original fluid mosaic model for biological membranes. Although their small size and dynamic nature have made them very difficult to study, it appears that lipid rafts may play a role in concentrating proteins that participate in signal transduction pathways and may also serve to regulate membrane curvature and budding.

All biological membranes are asymmetric

Membranes are structurally and functionally asymmetric. The outer and inner surfaces of *all known biological membranes have different components and different* *enzymatic activities.* A clear-cut example is the pump that regulates the concentration of Na^+ and K^+ ions in cells (Figure 12.33). This transport protein is located in the plasma membrane of nearly all cells in higher organisms. The Na^+–K^+ pump is oriented so that it pumps Na^+ out of the cell and K^+ into it. Furthermore, ATP must be on the inside of the cell to drive the pump. Ouabain, a specific inhibitor of the pump, is effective only if it is located outside. We shall consider the mechanism of this important and fascinating family of pumps in Chapter 13.

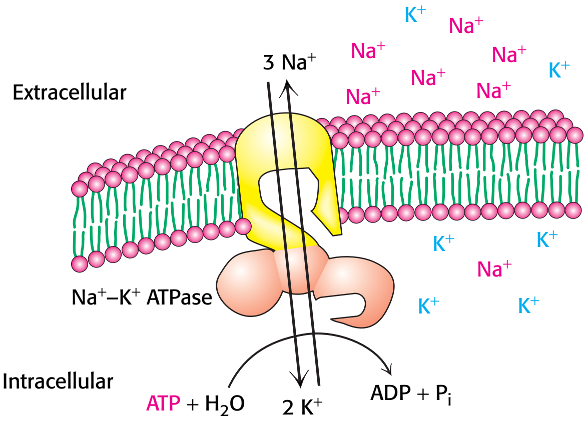


FIGURE 12.33 Asymmetry of the Na^+–K^+ transport system in plasma membranes. The Na^+–K^+ transport system pumps Na^+ out of the cell and K^+ into the cell by hydrolyzing ATP on the intracellular side of the membrane.

Membrane proteins have a unique orientation because, after synthesis, they are inserted into the membrane in an asymmetric manner. This absolute asymmetry is preserved because membrane proteins do not rotate from one side of the membrane to the other and because *membranes are always synthesized by the growth of preexisting membranes.* Lipids, too, are asymmetrically distributed as a consequence of their mode of biosynthesis, but this asymmetry is usually not absolute, except for glycolipids. In the red-blood-cell membrane, sphingomyelin and phosphatidylcholine are preferentially located in the outer leaflet of the bilayer, whereas phosphatidylethanolamine and phosphatidylserine are located mainly in the inner leaflet. Large amounts of cholesterol are present in both leaflets.

12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes

Thus far, we have considered only the plasma membrane of cells. Some bacteria and archaea have only this single membrane, surrounded by a cell wall. Other bacteria, such as *E. coli,* have two membranes separated by a cell wall (made of proteins, peptides, and carbohydrates) lying between them (Figure 12.34). The inner membrane acts as the permeability barrier, and the outer membrane and the cell wall provide additional protection. The outer membrane is quite permeable to small molecules, owing to the presence of porins. The region between the two membranes containing the cell wall is called the *periplasm.*

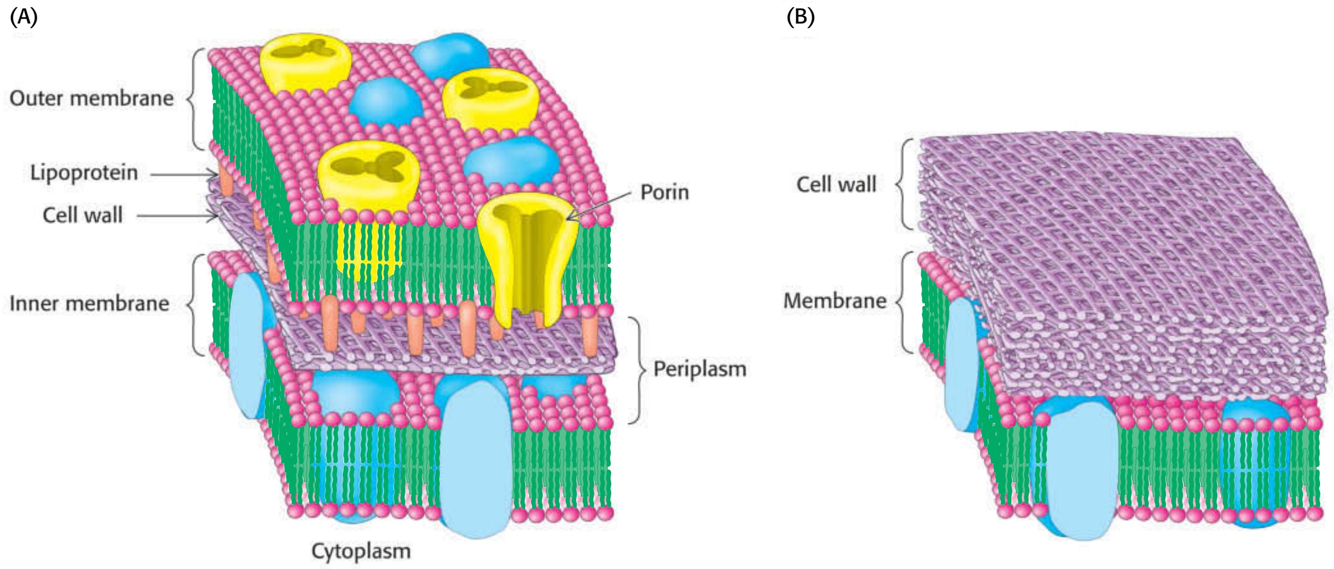


FIGURE 12.34 Cell membranes of prokaryotes. A schematic view of the membrane of bacterial cells surrounded by (A) two membranes or (B) one membrane.

Eukaryotic cells, with the exception of plant cells, do not have cell walls, and their cell membranes consist of a single lipid bilayer. In plant cells, the cell wall is on the outside of the plasma membrane. Eukaryotic cells are distinguished from prokaryotic cells by the presence of membranes inside the cell that form internal compartments. For example, peroxisomes, organelles that play a major role in the oxidation of fatty acids for energy conversion, are defined by a single membrane. Mitochondria, the organelles in which ATP is synthesized, are surrounded by two membranes. As in the case for a bacterium, the outer membrane is quite permeable to small molecules, whereas the inner membrane is not. Indeed, considerable evidence now indicates that mitochondria evolved from bacteria by *endosymbiosis* (Section 18.1). The nucleus is also surrounded by a double membrane, the *nuclear envelope*, that consists of a set of closed membranes that come together at structures called *nuclear pores* (Figure 12.35). These pores regulate transport into and out of the nucleus. The nuclear envelope is linked to another membrane-defined structure, the *endoplasmic reticulum*, which plays a host of cellular roles, including drug detoxification and the modification of proteins for secretion. Thus, a eukaryotic cell contains interacting compartments, and transport into and out of these compartments is essential to many biochemical processes.

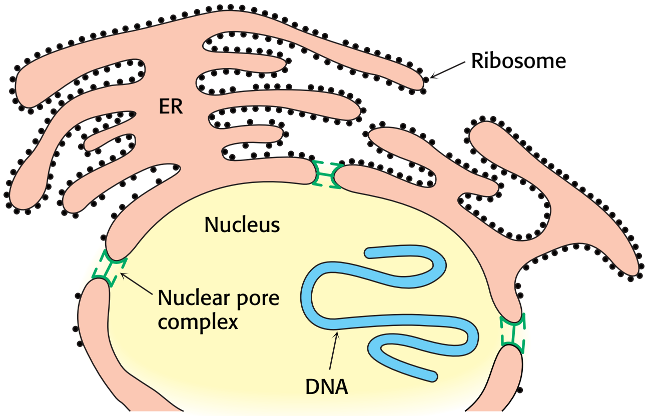


FIGURE 12.35 Nuclear envelope. The nuclear envelope is a double membrane connected to another membrane system of eukaryotes, the endoplasmic reticulum. [Information from E. C. Schirmer and L. Gerace, *Genome Biol*. 3(4):1008.1–1008.4, 2002, reviews, Fig.1.]

Membranes must be able to separate or join together so that cells and compartments may take up, transport, and release molecules. Many cells take up molecules through the process of *receptor-mediated endocytosis*. Here, a protein or larger complex initially binds to a receptor on the cell surface. After the receptor is bound, specialized proteins act to cause the membrane in this region to invaginate. One of these specialized proteins is *clathrin*, which polymerizes into a lattice network around the growing membrane bud, often referred to as a *clathrin-coated pit* (Figure 12.36). The invaginated membrane eventually breaks off and fuses to form a *vesicle*. Various hormones, transport proteins, and antibodies employ receptor-mediated endocytosis to gain entry into a cell. A less-advantageous consequence is that this pathway is available to viruses and toxins as a means of invading cells. The reverse process—the fusion of a vesicle to a membrane—is a key step in the release of neurotransmitters from a neuron into the synaptic cleft (Figure 12.37).

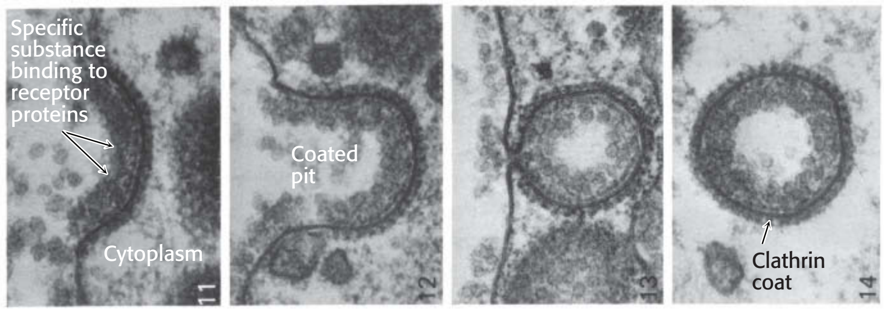


FIGURE 12.36 Vesicle formation by receptor-mediated endocytosis. Receptor binding on the surface of the cell induces the membrane to invaginate, with the assistance of specialized intracellular proteins such as clathrin. The process results in the formation of a vesicle within the cell. [M. M. Perry and A. B. Gilbert, *J. Cell Sci*. 39:266, 1979.]

Let us consider one example of receptor-mediated endocytosis. Iron is a critical element for the function and structure of many proteins, including hemoglobin and myoglobin (Chapter 7). However, free iron ions are highly toxic to cells, owing to their ability to catalyze the formation of free radicals. Hence, the transport of iron atoms from the digestive tract to the cells where they are most needed must be tightly controlled. In the bloodstream, iron is bound very tightly by the protein *transferrin*, which can bind two Fe^{3+} ions with a dissociation constant of 10^{-23} M at neutral pH. Cells requiring iron express the *transferrin receptor* in their plasma membranes (Section 32.4). Formation of a complex between the transferrin receptor and iron-bound transferrin initiates receptor-mediated endocytosis, internalizing these complexes within vesicles called *endosomes* (Figure 12.38). As the endosomes mature, proton pumps within the vesicle membrane lower the lumenal pH to about 5.5. Under these conditions, the affinity of iron ions for transferrin is reduced; these ions are released and are free to pass through channels in the endosomal membranes into the cytoplasm. The iron-free transferrin complex is recycled to the plasma membrane, where transferrin is released back into the bloodstream and the transferrin receptor can participate in another uptake cycle.

Although budding and fusion appear deceptively simple, the structures of the intermediates in these processes and the detailed mechanisms remain ongoing areas of investigation. These processes must be highly specific in order to prevent incorrect membrane fusion events. *SNARE* (soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor) *proteins* from both membranes help draw appropriate lipid bilayers together by forming tightly coiled four-helical bundles (Figure 12.39). Once these membranes are in close apposition, the fusion process can proceed. SNARE proteins, encoded by gene families in all eukaryotic cells, largely determine the compartment with which a vesicle will fuse. The specificity of membrane fusion ensures the orderly trafficking of membrane vesicles and their cargos through eukaryotic cells.

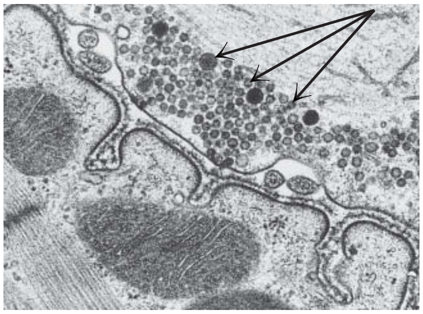


FIGURE 12.37 Neurotransmitter release. Neurotransmitter-containing synaptic vesicles (indicated by the arrows) are arrayed near the plasma membrane of a nerve cell. Synaptic vesicles fuse with the plasma membrane, releasing the neurotransmitter into the synaptic cleft. [T. Reese/ Don Fawcett/Photo Researchers.]

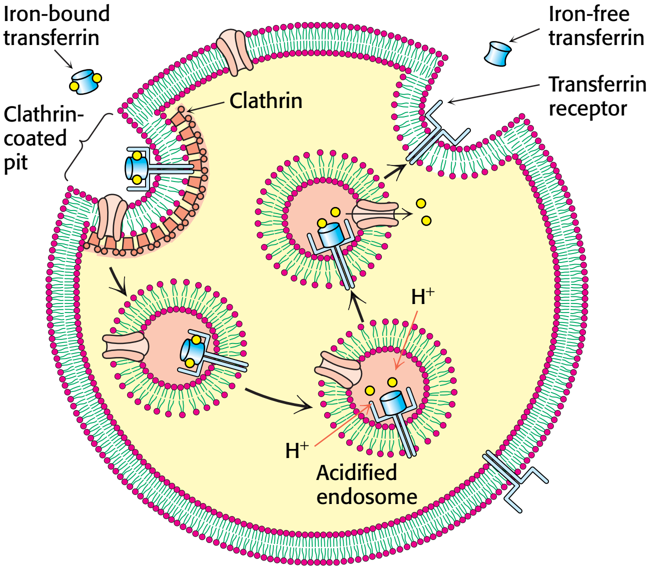


FIGURE 12.38 The transferrin receptor cycle. Iron-bound transferrin binds to the transferrin receptor (TfR) on the surface of cells. Receptor-mediated endocytosis occurs, leading to the formation of a vesicle called an endosome. As the lumen of the endosome is acidified by the action of proton pumps, iron is released from transferrin, passes through channels in the membrane, and is utilized by the cell. The complex between iron-free transferrin and the transferrin receptor is returned to the plasma membrane for another cycle. [Information from L. Zecca et al., *Nat. Rev. Neurosci*. 5:863–873, 2004, Fig.1.]

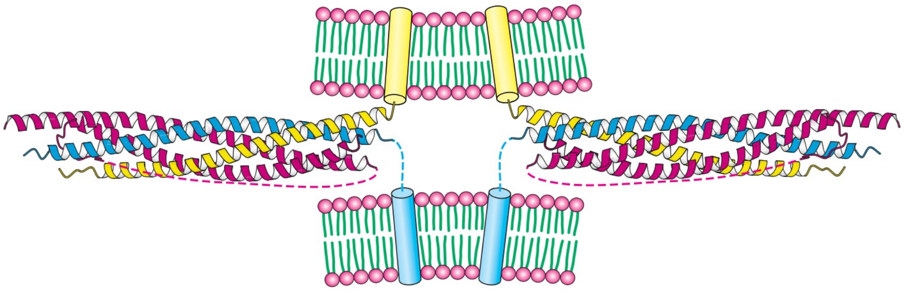


FIGURE 12.39 SNARE complexes initiate membrane fusion. The SNARE protein synaptobrevin (yellow) from one membrane forms a tight four-helical bundle with the corresponding SNARE proteins syntaxin-1 (blue) and SNAP25 (red) from a second membrane. The complex brings the membranes close together, initiating the fusion event. [Drawn from 1SFC.pdb.]

SUMMARY

Biological membranes are sheetlike structures, typically from 60 to 100 Å thick, that are composed of protein and lipid molecules held together by noncovalent interactions. Membranes are highly selective permeability barriers. They create closed compartments, which may be entire cells or organelles within a cell. Proteins in membranes regulate the molecular and ionic compositions of these compartments. Membranes also control the flow of information between cells.

12.1 Fatty Acids Are Key Constituents of Lipids

Fatty acids are hydrocarbon chains of various lengths and degrees of unsaturation that terminate with a carboxylic acid group. The fatty acid chains in membranes usually contain between 14 and 24 carbon atoms; they may be saturated or unsaturated. Short chain length and unsaturation enhance the fluidity of fatty acids and their derivatives by lowering the melting temperature.

12.2 There Are Three Common Types of Membrane Lipids

The major types of membrane lipids are phospholipids, glycolipids, and cholesterol. Phosphoglycerides, a type of phospholipid, consist of a glycerol backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine are major phosphoglycerides. Sphingomyelin, a different type of phospholipid, contains a sphingosine backbone instead of glycerol. Glycolipids are sugar-containing lipids derived from sphingosine. Cholesterol, which modulates membrane fluidity, is constructed from a steroid nucleus. A common feature of these membrane lipids is that they are amphipathic molecules, having one hydrophobic and one hydrophilic end.

12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media

Membrane lipids spontaneously form extensive bimolecular sheets in aqueous solutions. The driving force for membrane formation is the hydrophobic interactions among the fatty acid tails of membrane lipids. The hydrophilic head groups interact with the aqueous medium. Lipid bilayers are cooperative structures, held together by many weak bonds. These lipid bilayers are highly impermeable to ions and most polar molecules, yet they are quite fluid, which enables them to act as a solvent for membrane proteins.

12.4 Proteins Carry Out Most Membrane Processes

Specific proteins mediate distinctive membrane functions such as transport, communication, and energy transduction. Many integral membrane proteins span the lipid bilayer, whereas others are only partly embedded in the membrane. Peripheral membrane proteins are bound to membrane surfaces by electrostatic and hydrogen-bond interactions. Membrane-spanning proteins have regular structures, including β strands, although the α helix is the most common membrane-spanning structure. Sequences of 20 consecutive nonpolar amino acids can be diagnostic of a membrane-spanning α helical region of a protein.

12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane

Membranes are structurally and functionally asymmetric, as exemplified by the restriction of sugar residues to the external surface of mammalian plasma membranes. Membranes are dynamic structures in which proteins and lipids diffuse rapidly in the plane of the membrane (lateral diffusion), unless restricted by special interactions. In contrast, the rotation of lipids from one face of a membrane to the other (transverse diffusion, or flip-flop) is usually very slow. Proteins do not rotate across bilayers; hence, membrane asymmetry can be preserved. The degree of fluidity of a membrane depends on the chain length of its lipids and on the extent to which their constituent fatty acids are unsaturated. In animals, cholesterol content also regulates membrane fluidity.

12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes

An extensive array of internal membranes in eukaryotes creates compartments within a cell for distinct biochemical functions. For instance, a double membrane surrounds the nucleus (the location of most of the cell's genetic material) and the mitochondria (the location of most ATP synthesis). A single membrane defines the other internal compartments, such as the endoplasmic reticulum. Receptor-mediated endocytosis enables the formation of intracellular vesicles when ligands bind to their corresponding receptor proteins in the plasma membrane. The reverse process—the fusion of a vesicle to a membrane—is a key step in the release of signaling molecules outside the cell.

KEY TERMS

fatty acid (p. 342)
phospholipid (p. 344)
sphingosine (p. 344)
phosphoglyceride (p. 344)
sphingomyelin (p. 345)
glycolipid (p. 345)
cerebroside (p. 346)
ganglioside (p. 346)
cholesterol (p. 346)
amphipathic (amphiphilic) molecule (p. 347)
lipid bilayer (p. 348)
liposome (p. 348)
integral membrane protein (p. 351)
peripheral membrane protein (p. 351)
hydropathy plot (p. 355)
lateral diffusion (p. 356)
fluid mosaic model (p. 357)
lipid raft (p. 358)
receptor-mediated endocytosis (p. 360)
clathrin (p. 360)
transferrin (p. 361)
transferrin receptor (p. 361)
endosome (p. 361)
SNARE (soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor) proteins (p. 361)

PROBLEMS

1. *Population density.* How many phospholipid molecules are there in a 1-μm^2 region of a phospholipid bilayer membrane? Assume that a phospholipid molecule occupies 70 $Å^2$ of the surface area.

2. *Through the looking-glass.* Phospholipids form lipid bilayers in water. What structure might form if phospholipids were placed in an organic solvent?

3. *Lipid diffusion.* What is the average distance traversed by a membrane lipid in 1 μs, 1 ms, and 1 s? Assume a diffusion coefficient of 10^{-8} cm^2 s^{-1}.

4. *Protein diffusion.* The diffusion coefficient, *D*, of a rigid spherical molecule is given by

$$D = kT/6\pi\eta r$$

in which η is the viscosity of the solvent, *r* is the radius of the sphere, *k* is the Boltzman constant (1.38×10^{-16} erg degree^{-1}), and *T* is the absolute temperature. What is the diffusion coefficient at 37°C of a 100-kDa protein in a membrane that has an effective viscosity of 1 poise (1 poise = 1 erg s^{-1} cm^{-3})? What is the average distance traversed by this protein in 1 μs, 1 ms, and 1 s? Assume that this protein is an unhydrated, rigid sphere of density 1.35 g cm^{-3}.

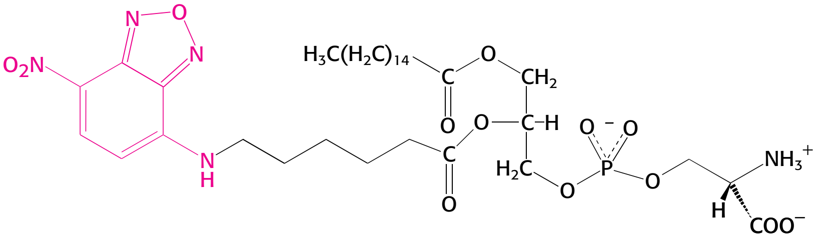
5. *Cold sensitivity.* Some antibiotics act as carriers that bind an ion on one side of a membrane, diffuse through the membrane, and release the ion on the other side. The conductance of a lipid-bilayer membrane containing a carrier antibiotic decreased abruptly when the temperature was lowered from 40°C to 36°C. In contrast, there was little change in conductance of the same bilayer membrane when it contained a channel-forming antibiotic. Why?

6. *Melting point 1.* Explain why oleic acid (18 carbons, one cis bond) has a lower melting point than stearic acid, which has the same number of carbon atoms but is saturated. How would you expect the melting point of *trans*-oleic acid to compare with that of *cis*-oleic acid? Why might most unsaturated fatty acids in phospholipids be in the cis rather than the trans conformation?

7. *Melting point 2.* Explain why the melting point of palmitic acid (C_{16}) is 6.5 degrees lower than that of stearic acid (C_{18}).

8. *A sound diet.* Small mammalian hibernators can withstand body temperatures of 0° to 5°C without injury. However, the body fats of most mammals have melting temperatures of approximately 25°C. Predict how the composition of the body fat of hibernators might differ from that of their nonhibernating cousins.

9. *Flip-flop 1.* The transverse diffusion of phospholipids in a bilayer membrane was investigated by using a fluorescently labeled analog of phosphatidylserine called NBD-PS.



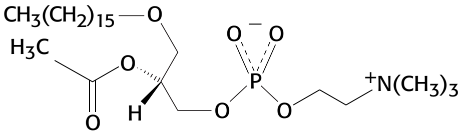
NBD-phosphatidylserine (NBD-PS)

The fluorescence signal of NBD-PS is quenched when exposed to sodium dithionite, a reducing agent that is not membrane permeable.

Lipid vesicles containing phosphatidylserine (98%) and NBD-PS (2%) were prepared by sonication and purified. Within a few minutes of the addition of sodium dithionite, the fluorescence signal of these vesicles decreased to ~45% of its initial value. Immediately adding a second addition of sodium dithionite yielded no change in the fluorescence signal. However, if the vesicles were allowed to incubate for 6.5 hours, a third addition of sodium dithionite decreased the remaining fluorescence signal by 50%. How would you interpret the fluorescence changes at each addition of sodium dithionite?

10. *Flip-flop 2.* Although proteins rarely if ever flip-flop across a membrane, the distribution of membrane lipids between the membrane leaflets is not absolute except for glycolipids. Why are glycosylated lipids less likely to flip-flop?

11. *Linkages.* Platelet-activating factor (PAF) is a phospholipid that plays a role in allergic and inflammatory responses, as well as in toxic shock syndrome. The structure of PAF is shown here. How does it differ from the structures of the phospholipids discussed in this chapter?



Platelet-activating factor (PAF)

12. *A question of competition.* Would a homopolymer of alanine be more likely to form an α helix in water or in a hydrophobic medium? Explain.

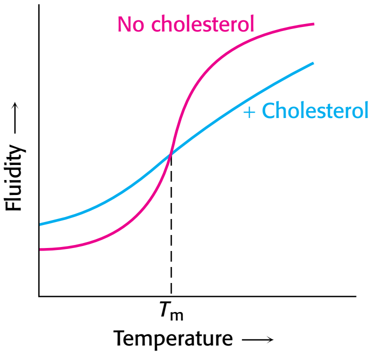
13. *A false positive.* Hydropathy plot analysis of your protein of interest reveals a single, prominent hydrophobic peak. However, you later discover that this protein is soluble and not membrane associated. Explain how the hydropathy plot may have been misleading.

14. *Maintaining fluidity.* A culture of bacteria growing at 37°C was shifted to 25°C. How would you expect this shift to alter the fatty acid composition of the membrane phospholipids? Explain.

15. *Let me count the ways.* Each intracellular fusion of a vesicle with a membrane requires a SNARE protein on the vesicle (called the v-SNARE) and a SNARE protein on the target membrane (called the t-SNARE). Assume that a genome encodes 21 members of the v-SNARE family and 7 members of the t-SNARE family. With the assumption of no specificity, how many potential v-SNARE–t-SNARE interactions could take place?

Data Interpretation Problems

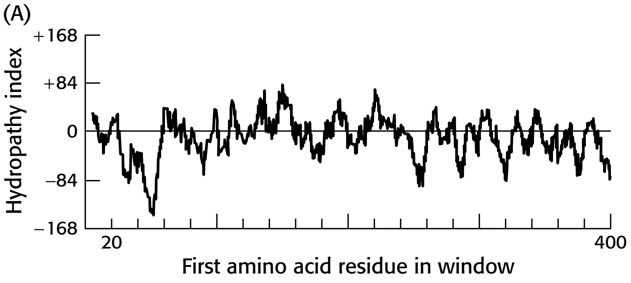
16. *Cholesterol effects.* The red curve on the following graph shows the fluidity of the fatty acids of a phospholipid bilayer as a function of temperature. The blue curve shows the fluidity in the presence of cholesterol.

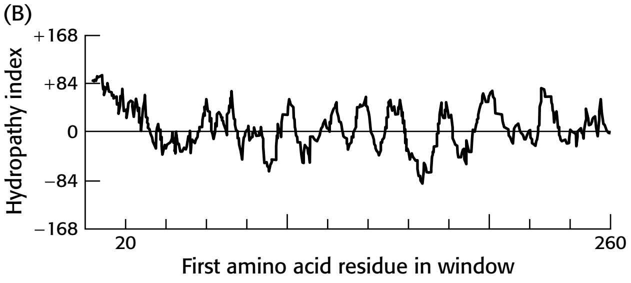


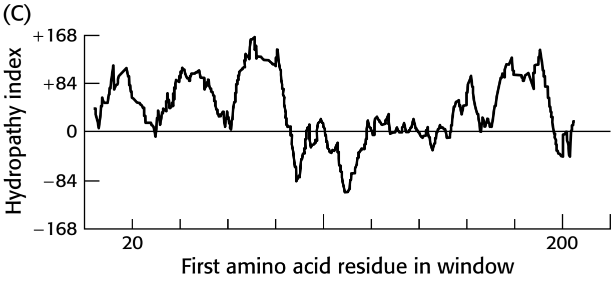
(a) What is the effect of cholesterol?

(b) Why might this effect be biologically important?

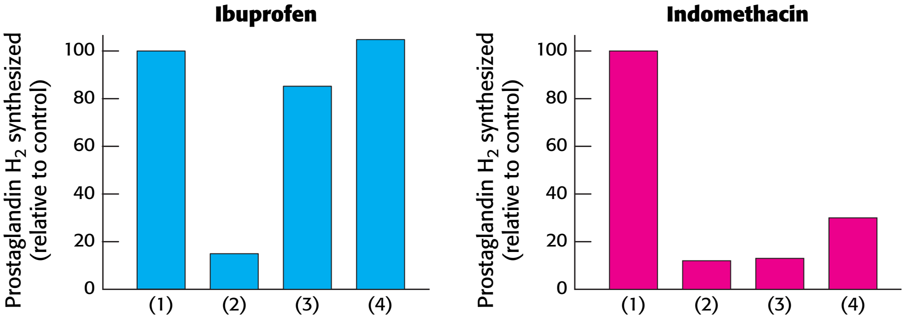
17. *Hydropathy plots.* On the basis of the following hydropathy plots for three proteins (A–C), predict which would be membrane proteins. What are the ambiguities with respect to using such plots to determine if a protein is a membrane protein?







18. *Not all inhibitors are equal.* Ibuprofen and indomethacin are clinically important inhibitors of prostaglandin H_2 synthase-1. Cells expressing this enzyme were incubated under the following conditions, after which the activity of the enzyme was measured by adding radiolabeled arachidonic acid and detecting newly-produced prostaglandin H_2:



(1) 40 min *without* inhibitor (control)

(2) 40 min with inhibitor

(3) 40 min with inhibitor, after which the cells were resuspended in medium *without* inhibitor

(4) 40 min with inhibitor, after which the cells were resuspended in medium *without* inhibitor and incubated for an additional 30 min.

(a) Provide a hypothesis explaining the different results for these two inhibitors.

(b) How would these results look if aspirin were tested in a similar fashion?

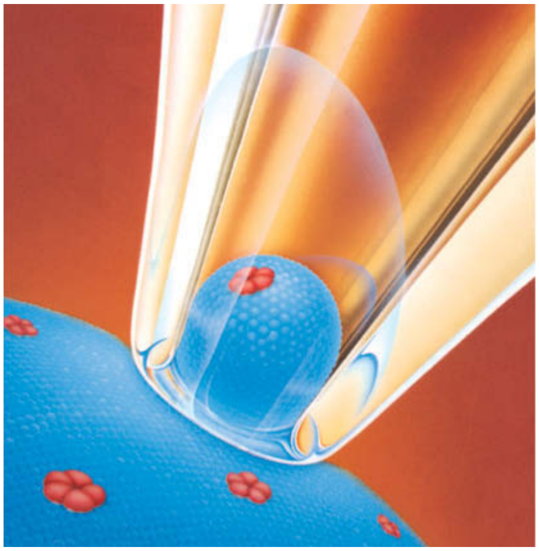
Chapter Integration Problem

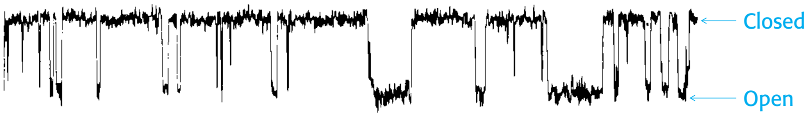
19. *The proper environment.* An understanding of the structure and function of membrane proteins has lagged behind that of other proteins. The primary reason is that membrane proteins are more difficult to purify and crystallize. Why might this be the case?

CHAPTER

13

Membrane Channels and Pumps





The flow of ions through a single membrane channel (channels are shown in red in the illustration at the left) can be detected by the patch-clamp technique, which records current changes as the channel alternates between open and closed states. [(Left) Information from E. Neher and B. Sakmann. The patch clamp technique. Copyright © 1992 by Scientific American, Inc. All rights reserved. (Right) Courtesy of Dr. Mauricio Montal.]

The lipid bilayer of biological membranes is intrinsically impermeable to ions and polar molecules, yet these species must be able to cross these membranes for normal cell function. Permeability is conferred by three classes of membrane proteins: *pumps, carriers,* and *channels.* Pumps use a source of free energy such as ATP hydrolysis or light absorption to drive the thermodynamically uphill transport of ions or molecules. Pump action is an example of *active transport*. Carriers mediate the transport of ions and small molecules across the membrane without consumption of ATP. Channels provide a membrane pore through which ions can flow very rapidly in a thermodynamically downhill direction. The action of channels illustrates *passive transport*, or *facilitated diffusion*.

Pumps are energy transducers in that they convert one form of free energy into another. Two types of *ATP-driven pumps*, P-type ATPases and the ATP-binding cassette (ABC) transporters, undergo conformational changes on ATP binding and hydrolysis that cause a bound ion to be transported across the membrane. The free energy of ATP hydrolysis is used to drive the movement of ions against their concentration gradients, a process referred to as *primary active transport*. In contrast, carriers utilize the gradient of one ion to drive the transport of another against its gradient. An example of this process, termed *secondary active transport,* is mediated by the *E. coli* lactose transporter, a well-studied protein responsible for the uptake of a specific sugar from the environment of a bacterium. Many transporters of this class are present in the membranes of our cells. The expression of these transporters determines which metabolites a cell can import from the environment. Hence, adjusting the level of transporter expression is a primary means of controlling metabolism.

OUTLINE

13.1 The Transport of Molecules Across a Membrane May Be Active or Passive

13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions and Molecules Across Membranes

13.3 Lactose Permease Is an Archetype of Secondary Transporters That Use One Concentration Gradient to Power the Formation of Another

13.4 Specific Channels Can Rapidly Transport Ions Across Membranes

13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells

13.6 Specific Channels Increase the Permeability of Some Membranes to Water

Pumps can establish persistent gradients of particular ions across membranes. Specific *ion channels* can allow these ions to flow rapidly across membranes down these gradients. These channels are among the most fascinating molecules in biochemistry in their ability to allow some ions to flow freely through a membrane while blocking the flow of even closely related species. The opening, or gating, of these channels can be controlled by the presence of certain ligands or a particular membrane voltage. Gated ion channels are central to the functioning of our nervous systems, acting as elaborately switched wires that allow the rapid flow of current.

Finally, a different class of channel, the cell-to-cell channel, or *gap junction*, allows the flow of metabolites or ions *between cells*. For example, gap junctions are responsible for synchronizing muscle-cell contraction in the beating heart.

The expression of transporters largely defines the metabolic activities of a given cell type

Each cell type expresses a specific set of transporters in its plasma membrane. This collection of expressed transporters is important because it largely determines the ionic composition inside cells and the compounds that can be taken up from the extracellular environment. In some sense, the specific array of transporters defines the cell's characteristics because a cell can execute only those biochemical reactions for which it has taken up the necessary substrates.

An example from glucose metabolism illustrates this point. As we will see in Chapter 16, tissues differ in their ability to employ different molecules as energy sources. Which tissues can utilize glucose is largely governed by the expression of members of the GLUT family of homologous glucose transporters. For example, GLUT3 is the primary glucose transporter expressed on the plasma membrane of neurons. This transporter binds glucose relatively tightly so that these cells have first call on glucose when it is present at relatively low concentrations. We will encounter many such examples of the critical role that transporter expression plays in the control and integration of metabolism.

13.1 The Transport of Molecules Across a Membrane May Be Active or Passive

We first consider some general principles of membrane transport. Two factors determine whether a molecule will cross a membrane: (1) the permeability of the molecule in a lipid bilayer and (2) the availability of an energy source.

Many molecules require protein transporters to cross membranes

As stated in Chapter 12, some molecules can pass through cell membranes because they dissolve in the lipid bilayer. Such molecules are called *lipophilic molecules.* The steroid hormones provide a physiological example. These cholesterol relatives can pass through a membrane, but what determines the direction in which they will move? Such molecules will pass through a membrane down their concentration gradient in a process called *simple diffusion*. In accord with the Second Law of Thermodynamics, molecules spontaneously move from a region of higher concentration to one of lower concentration.

Matters become more complicated when the molecule is highly polar. For example, sodium ions are present at 143 mM outside a typical cell and at 14 mM inside the cell. However, sodium does not freely enter the cell, because the charged ion cannot pass through the hydrophobic membrane interior. In some circumstances, as during a nerve impulse, sodium ions must enter the cell. How are they able to do so? Sodium ions pass through specific channels in the hydrophobic barrier formed by membrane proteins. This means of crossing the membrane is called *facilitated diffusion* because the diffusion across the membrane is facilitated by the channel. It is also called *passive transport* because the energy driving the ion movement originates from the ion gradient itself, without any contribution by the transport system. Channels, like enzymes, display substrate specificity in that they facilitate the transport of some ions, but not other, even closely related, ions.

How is the sodium gradient established in the first place? In this case, sodium must move, or be pumped, *against* a concentration gradient. Because moving the ion from a low concentration to a higher concentration results in a decrease in entropy, it requires an input of free energy. Protein transporters embedded in the membrane are capable of using an energy source to move the molecule up a concentration gradient. Because an input of energy from another source is required, this means of crossing the membrane is called *active transport*.

Free energy stored in concentration gradients can be quantified

An unequal distribution of molecules is an energy-rich condition because free energy is minimized when all concentrations are equal. Consequently, to attain such an unequal distribution of molecules requires an input of free energy. How can we quantify the amount of energy required to generate a concentration gradient (Figure 13.1)? Consider an uncharged solute molecule. The free-energy change in transporting this species from side 1, where it is present at a concentration of c_1, to side 2, where it is present at concentration c_2, is

$$\Delta G = RT \ln(c_2/c_1)$$

where R is the gas constant (8.315×10^{-3} kJ mol^{-1} deg^{-1}, or 1.987×10^{-3} kcal mol^{-1} deg^{-1}) and T is the temperature in kelvins. For a charged species, the unequal distribution across the membrane generates an electrical potential that also must be considered because the ions will be repelled by the like charges. The sum of the concentration and electrical terms is called the *electrochemical potential* or *membrane potential*. The free-energy change is then given by

$$\Delta G = RT \ln(c_2/c_1) + ZF\Delta V$$

in which Z is the electrical charge of the transported species, ΔV is the potential in volts across the membrane, and F is the Faraday constant (96.5 kJ V^{-1} mol^{-1}, or 23.1 kcal V^{-1} mol^{-1}).

A transport process must be active when ΔG is positive, whereas it can be passive when ΔG is negative. For example, consider the transport of an uncharged molecule from $c_1 = 10^{-3}$ M to $c_2 = 10^{-1}$ M.

$$\begin{aligned}\Delta G &= RT \ln(10^{-1}/10^{-3}) \\ &= (8.315 \times 10^{-3}) \times 298 \times \ln(10^2) \\ &= +11.4 \text{ kJ mol}^{-1} (+2.7 \text{ kcal mol}^{-1})\end{aligned}$$

At 25°C (298 K), ΔG is +11.4 kJ mol^{-1} (+2.7 kcal mol^{-1}), indicating that this transport process requires an input of free energy.

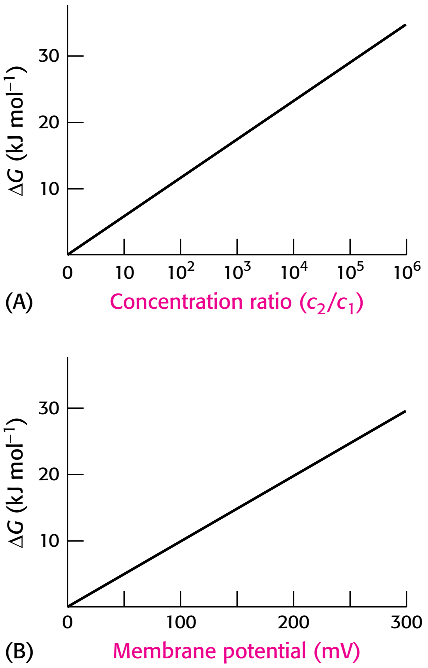


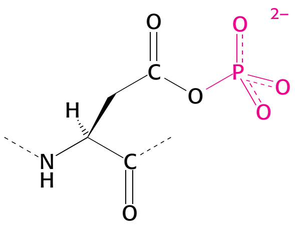
FIGURE 13.1 Free energy and transport. The free-energy change in transporting (A) an uncharged solute from a compartment at concentration c_1 to one at c_2 and (B) a singly charged species across a membrane to the side having the same charge as that of the transported ion. *Notice* that the free-energy change imposed by a membrane potential of 59 mV is equivalent to that imposed by a concentration ratio of 10 for a singly charged ion at 25°C.

13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions and Molecules Across Membranes

The extracellular fluid of animal cells has a salt concentration similar to that of seawater. However, cells must control their intracellular salt concentrations to facilitate specific processes, such as signal transduction and action potential propagation, and prevent unfavorable interactions with high concentrations of ions such as Ca^{2+}. For instance, most animal cells contain a high concentration of K^+ and a low concentration of Na^+ relative to the external medium. These ionic gradients are generated by a specific transport system, an enzyme that is called the *Na^+–K^+ pump* or the *Na^+–K^+ ATPase.* The hydrolysis of ATP by the pump provides the energy needed for the active transport of Na^+ out of the cell and K^+ into the cell, generating the gradients. The pump is called the Na^+–K^+ ATPase because the hydrolysis of ATP takes place only when Na^+ and K^+ are present. This ATPase, like all such enzymes, requires Mg^{2+}.

The change in free energy accompanying the transport of Na^+ and K^+ can be calculated. Suppose that the concentrations of Na^+ outside and inside the cell are 143 and 14 mM, respectively, and the corresponding values for K^+ are 4 and 157 mM. At a membrane potential of -50 mV and a temperature of 37°C, we can use the equation on page 369 to determine that the free-energy change in transporting 3 mol of Na^+ out of the cell and 2 mol of K^+ into the cell is $3(5.99) + 2(9.46) = +36.9$ kJ mol^{-1} ($+8.8$ kcal mol^{-1}). Under typical cellular conditions, the hydrolysis of a single ATP molecule per transport cycle provides sufficient free energy, about -50 kJ mol^{-1} (-12 kcal mol^{-1}) to drive the uphill transport of these ions. The active transport of Na^+ and K^+ is of great physiological significance. Indeed, more than a third of the ATP consumed by a resting animal is used to pump these ions. The Na^+–K^+ gradient in animal cells controls cell volume, renders neurons and muscle cells electrically excitable, and drives the active transport of sugars and amino acids.

The purification of other ion pumps has revealed a large family of evolutionarily related ion pumps including proteins from bacteria, archaea, and all eukaryotes. Each of these pumps is specific for a particular ion or set of ions. Two are of particular interest: the *sarcoplasmic reticulum Ca^{2+} ATPase* (or *SERCA*) transports Ca^{2+} out of the cytoplasm and into the sarcoplasmic reticulum of muscle cells, and the *gastric H^+–K^+ ATPase* is the enzyme responsible for pumping sufficient protons into the stomach to lower the pH to 1.0. These enzymes and the hundreds of known homologs, including the Na^+–K^+ ATPase, are referred to as *P-type ATPases* because they form a key phosphorylated intermediate. In the formation of this intermediate, a phosphoryl group from ATP is linked to the side chain of a specific conserved aspartate residue in the ATPase to form phosphorylaspartate.



Phosphorylaspartate

P-type ATPases couple phosphorylation and conformational changes to pump calcium ions across membranes

Membrane pumps function by mechanisms that are simple in principle but often complex in detail. Fundamentally, each pump protein can exist in two principal conformational states, one with ion-binding sites open to one side of the membrane and the other with ion-binding sites open to the other side (Figure 13.2). To pump ions in a single direction across a membrane, the free energy of ATP hydrolysis must be coupled to the interconversion between these conformational states.

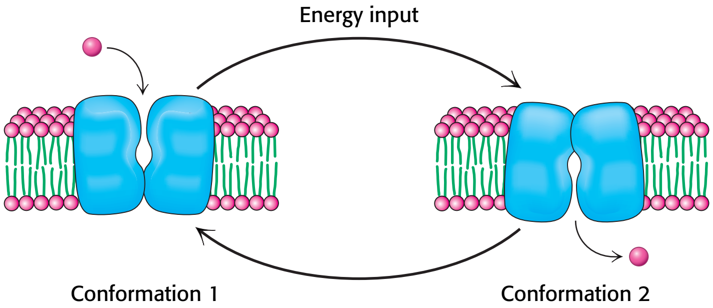


FIGURE 13.2 Pump action. A simple scheme for the pumping of a molecule across a membrane. The pump interconverts to two conformational states, each with a binding site accessible to a different side of the membrane.

We will consider the structural and mechanistic features of P-type ATPases by examining SERCA. The properties of this P-type ATPase have been established in great detail by relying on crystal structures of the pump in five different states. This enzyme, which constitutes 80% of the protein in the sarcoplasmic reticulum membrane, plays an important role in relaxation of contracted muscle. Muscle contraction is triggered by an abrupt rise in the cytoplasmic calcium ion level. Subsequent muscle relaxation depends on the rapid removal of Ca^{2+} from the cytoplasm into the sarcoplasmic reticulum, a specialized compartment for Ca^{2+} storage, by SERCA. This pump maintains a Ca^{2+} concentration of approximately 0.1 μM in the cytoplasm compared with 1.5 mM in the sarcoplasmic reticulum.

The first structure of SERCA to be determined had Ca^{2+} bound, but no nucleotides present (Figure 13.3). SERCA is a single 110-kDa polypeptide with a transmembrane domain consisting of 10 α helices. The transmembrane domain includes sites for binding two calcium ions. Each calcium ion is coordinated to seven oxygen atoms coming from a combination of side-chain glutamate, aspartate, threonine, and asparagine residues, backbone carbonyl groups, and water molecules. A large cytoplasmic headpiece constitutes nearly half the molecular weight of the protein and consists of three distinct domains, each with a distinct function. One domain (N) binds the ATP *n*ucleotide, another (P) accepts the *p*hosphoryl group on a conserved aspartate residue, and the third (A) serves as an

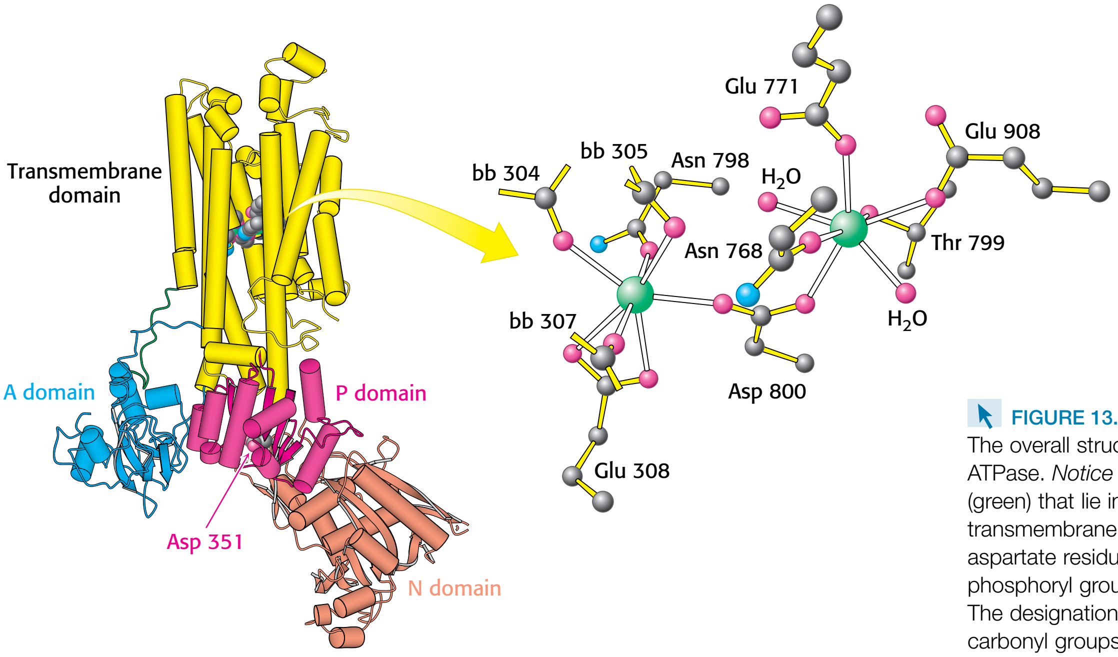


FIGURE 13.3 Calcium-pump structure. The overall structure of the SERCA P-type ATPase. *Notice* the two calcium ions (green) that lie in the center of the transmembrane domain. A conserved aspartate residue (Asp 351) that binds a phosphoryl group lies in the P domain. The designation bb refers to backbone carbonyl groups. [Drawn from 1SU4.pdb.]

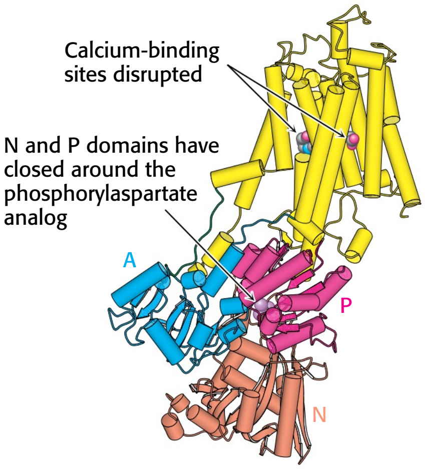

Calcium-binding sites disrupted
N and P domains have closed around the phosphorylaspartate analog
A
P
N

FIGURE 13.4 Conformational changes associated with calcium pumping. This structure was determined in the absence of bound calcium but with a phosphorylaspartate analog present in the P domain. *Notice* how different this structure is from the calcium-bound form shown in Figure 13.3: both the transmembrane part (yellow) and the A, P, and N domains have substantially rearranged. [Drawn from 1WPG.pdb.]

*a*ctuator, linking changes in the N and P domains to the transmembrane part of the enzyme.

SERCA is a remarkably dynamic protein. For example, the structure of SERCA without bound Ca^{2+}, but with a phosphorylaspartate analog present in the P domain, is shown in Figure 13.4. The N and P domains are now closed around the phosphorylaspartate analog, and the A domain has rotated substantially relative to its position in SERCA with Ca^{2+} bound and without the phosphoryl analog. Furthermore, the transmembrane part of the enzyme has rearranged significantly and the well-organized Ca^{2+}-binding sites are disrupted. These sites are now accessible from the side of the membrane opposite the N, P, and A domains.

The structural results can be combined with other studies to construct a detailed mechanism for Ca^{2+} pumping by SERCA (Figure 13.5):

1. The catalytic cycle begins with the enzyme in its unphosphorylated state with two calcium ions bound. We will refer to the overall enzyme conformation in this state as E_1; with Ca^{2+} bound, it is E_1-$(Ca^{2+})_2$. In this conformation, SERCA can bind calcium ions only on the cytoplasmic side of the membrane. This conformation is shown in Figure 13.3.

2. In the E_1 conformation, the enzyme can bind ATP. The N, P, and A domains undergo substantial rearrangement as they close around the bound ATP, but there is no substantial conformational change in the transmembrane domain. The calcium ions are now trapped inside the enzyme.

3. The phosphoryl group is then transferred from ATP to Asp 351.

4. Upon ADP release, the enzyme again changes its overall conformation, including the membrane domain this time. This new conformation is referred to as E_2 or E_2-P in its phosphorylated form. The process of interconverting the E_1 and E_2 conformations is sometimes referred to as *eversion*.

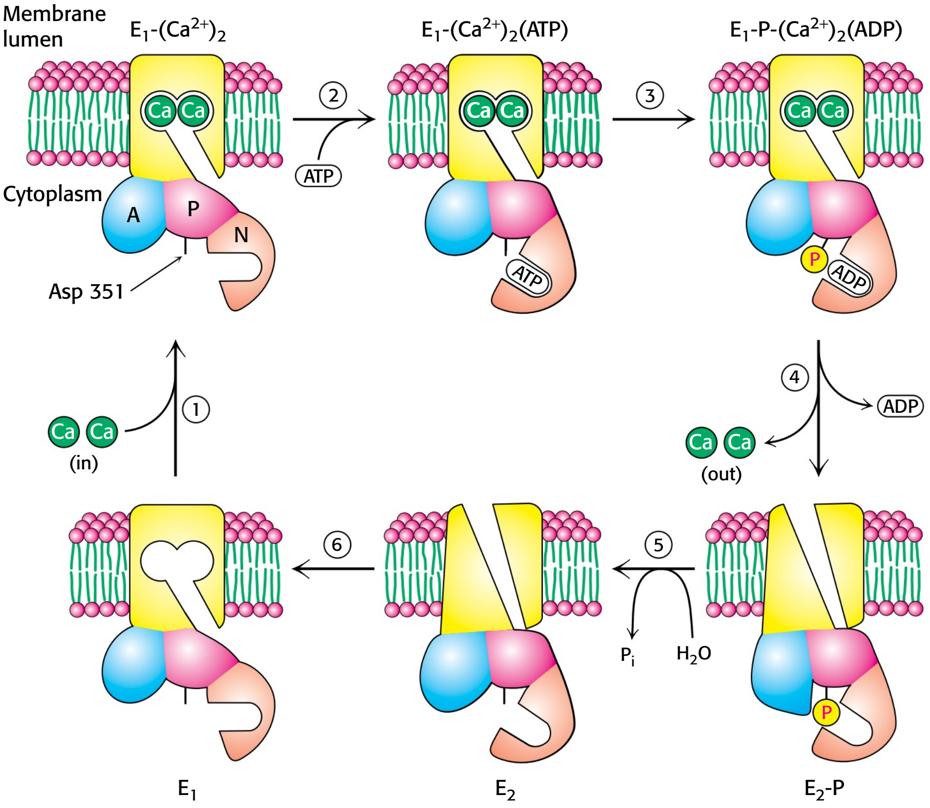


FIGURE 13.5 Pumping calcium. Ca^{2+}ATPase transports Ca^{2+} through the membrane by a mechanism that includes (1) Ca^{2+} binding from the cytoplasm, (2) ATP binding, (3) ATP cleavage with the transfer of a phosphoryl group to Asp 351 on the enzyme, (4) ADP release and eversion of the enzyme to release Ca^{2+} on the opposite side of the membrane, (5) hydrolysis of the phosphorylaspartate residue, and (6) eversion to prepare for the binding of Ca^{2+} from the cytoplasm.

In the E_2-P conformation, the Ca^{2+}-binding sites become disrupted and the calcium ions are released to the side of the membrane opposite that at which they entered; ion transport has been achieved. This conformation is shown in Figure 13.4.

5. The phosphorylaspartate residue is hydrolyzed to release inorganic phosphate.

6. With the release of phosphate, the interactions stabilizing the E_2 conformation are lost, and the enzyme everts to the E_1 conformation.

The binding of two calcium ions from the cytoplasmic side of the membrane completes the cycle.

This mechanism likely applies to other P-type ATPases. For example, Na^+–K^+ ATPase is an $\alpha_2\beta_2$ tetramer. Its α subunit is homologous to SERCA and includes a key aspartate residue analogous to Asp 351. The β subunit does not directly take part in ion transport. A mechanism analogous to that shown in Figure 13.5 applies, with three Na^+ ions binding from the inside of the cell to the E_1 conformation and two K^+ ions binding from outside the cell to the E_2 conformation.

Digitalis specifically inhibits the Na^+–K^+ pump by blocking its dephosphorylation

Certain steroids derived from plants are potent inhibitors ($K_i \approx 10$ nM) of the Na^+–K^+ pump. Digitoxigenin and ouabain are members of this class of inhibitors, which are known as *cardiotonic steroids* because of their strong effects on the heart (Figure 13.6). These compounds inhibit the dephosphorylation of the E_2-P form of the ATPase when applied on the *extracellular* face of the membrane.

Digitalis is a mixture of cardiotonic steroids derived from the dried leaf of the foxglove plant (*Digitalis purpurea*). The compound increases the force of contraction of heart muscle and is consequently a choice drug in the treatment of congestive heart failure. Inhibition of the Na^+–K^+ pump by digitalis leads to a higher level of Na^+ inside the cell. The diminished Na^+ gradient results in slower extrusion of Ca^{2+} by the sodium–calcium exchanger, an antiporter (Section 13.3). The subsequent increase in the intracellular level of Ca^{2+} enhances the ability of cardiac muscle to contract. It is interesting to note that digitalis was used effectively long before the discovery of the Na^+–K^+ ATPase. In 1785, William Withering, a British physician, heard tales of an elderly woman, known as "the old woman of Shropshire," who cured people of "dropsy" (which today would be recognized as congestive heart failure) with an extract of foxglove. Withering conducted the first scientific study of the effects of foxglove on congestive heart failure and documented its effectiveness.

Foxglove (*Digitalis purpurea*) is the source of digitalis, one of the most widely used drugs. [Roger Hall/Shutterstock.]

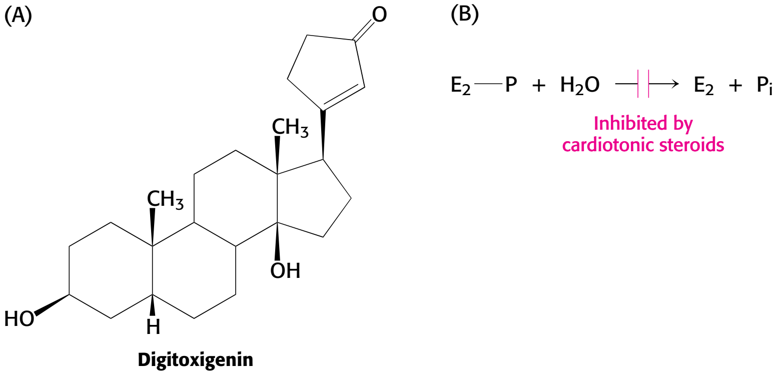


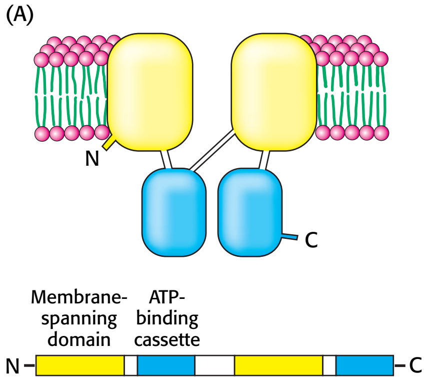
FIGURE 13.6 Digitoxigenin. Cardiotonic steroids such as digitoxigenin inhibit the Na^+–K^+ pump by blocking the dephosphorylation of E_2-P.

P-type ATPases are evolutionarily conserved and play a wide range of roles

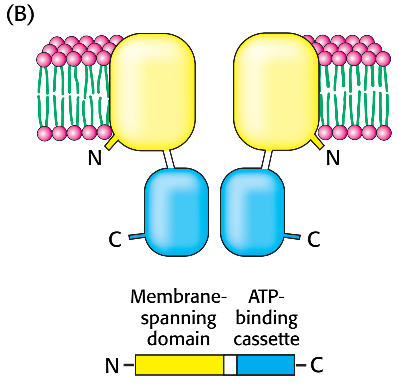
Analysis of the complete yeast genome revealed the presence of 16 proteins that clearly belong to the P-type ATPase family. More-detailed sequence analysis suggests that 2 of these proteins transport H^+ ions, 2 transport Ca^{2+}, 3 transport Na^+, and 2 transport metals such as Cu^{2+}. In addition, 5 members of this family appear to participate in the transport of phospholipids with amino acid head groups. These 5 proteins help maintain membrane asymmetry by transporting lipids such as phosphatidylserine from the outer to the inner leaflet of the bilayer membrane. Such enzymes have been termed "flippases." Remarkably, the human genome encodes 70 P-type ATPases. All members of this protein family employ the same fundamental mechanism: the free energy of ATP hydrolysis drives membrane transport by means of conformational changes, which are induced by the addition and removal of a phosphoryl group at an analogous aspartate site in each protein.

Multidrug resistance highlights a family of membrane pumps with ATP-binding cassette domains

Studies of human disease revealed another large and important family of active-transport proteins, with structures and mechanisms quite different from those of the P-type ATPase family. These pumps were identified from studies on tumor cells in culture that developed resistance to drugs that had been initially quite toxic to the cells. Remarkably, the development of resistance to one drug had made the cells less sensitive to a range of other compounds. This phenomenon is known as *multidrug resistance*. In a significant discovery, the onset of multidrug resistance was found to correlate with the expression and activity of a membrane protein with an apparent molecular mass of 170 kDa. This protein acts as an ATP-dependent pump that extrudes a wide range of small molecules from cells that express it. The protein is called the *multidrug-resistance* (MDR) *protein* or *P-glycoprotein* ("glyco" because it includes a carbohydrate moiety). Thus, when cells are exposed to a drug, the MDR pumps the drug out of the cell before the drug can exert its effects.



Multidrug-resistance protein (MDR)



Bacterial lipid transporter (MsbA)

FIGURE 13.7 Domain arrangement of ABC transporters. ABC transporters are a large family of homologous proteins composed of two transmembrane domains and two ATP-binding domains called ATP-binding cassettes (ABCs). (A) The multidrug-resistance protein is a single polypeptide chain containing all four domains, whereas (B) the bacterial lipid transporter MsbA consists of a dimer of two identical chains, containing one of each domain.

Analysis of the amino acid sequences of MDR and homologous proteins revealed a common architecture (Figure 13.7A). Each protein comprises four domains: two membrane-spanning domains and two ATP-binding domains. The ATP-binding domains of these proteins are called *ATP-binding cassettes* (ABCs) and are homologous to domains in a large family of transport proteins in bacteria and archaea. Transporters that include these domains are called *ABC transporters*. With 79 members, the ABC transporters are the largest single family identified in the *E. coli* genome. The human genome includes more than 150 ABC transporter genes.

The ABC proteins are members of the P-loop NTPase superfamily (Section 9.4). The three-dimensional structures of several members of the ABC transporter family have now been determined, including that of the bacterial lipid transporter MsbA. In contrast with the eukaryotic MDR protein, this protein is a dimer of 62-kDa chains: the amino-terminal half of each protein contains the membrane-spanning domain, and the carboxyl-terminal half contains the ATP-binding cassette (Figure 13.7B). Prokaryotic ABC proteins are often made up of multiple subunits, such as a dimer of identical chains, as above, or as a heterotetramer of two membrane-spanning domain subunits and two ATP-binding-cassette subunits. The consolidation of the enzymatic activities of several polypeptide chains in prokaryotes to a single chain in eukaryotes is a theme that we will see again. The two

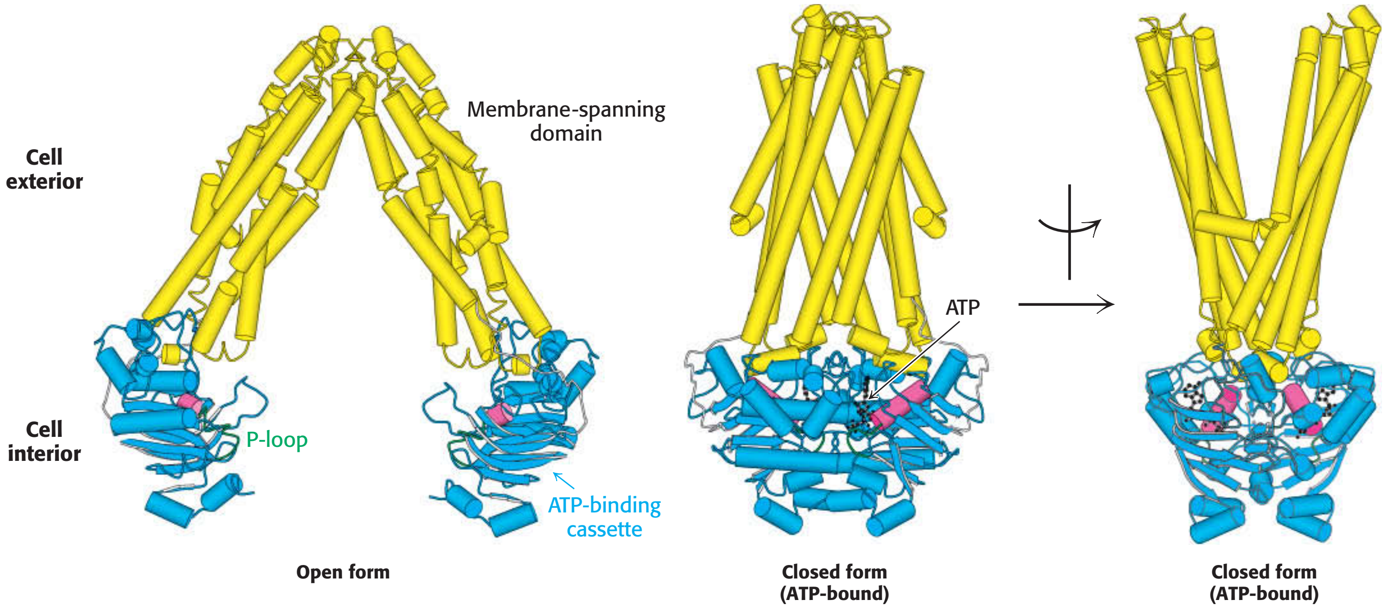

Cell exterior
Membrane-spanning domain
Cell interior
P-loop
ATP-binding cassette
ATP
Open form
Closed form (ATP-bound)
Closed form (ATP-bound)

FIGURE 13.8 ABC transporter structure. Two structures of the bacterial lipid transporter MsbA, a representative ABC transporter. The nucleotide-free, inward-facing form is on the left and the ATP-bound, outward-facing form is shown in two views (rotated by 90 degrees) in the center and on the right. The two ATP-binding cassettes (blue) are related to the P-loop NTPases and, like them, contain P-loops (green). The α helix adjacent to the P-loop is shown in red. [Drawn from 3B5W and 3B60.pdb.]

ATP-binding cassettes are in contact, but they do not interact strongly in the absence of bound ATP (Figure 13.8). On the basis of this structure and others, as well as on other experiments, a mechanism for active transport by these proteins has been developed (Figure 13.9):

1. The catalytic cycle begins with the transporter free of both ATP and substrate. While the distance between the ATP-binding cassettes in this form may vary with the individual transporter, the substrate binding region of the transporter faces inward.

2. Substrate enters the central cavity of the transporter from inside the cell. Substrate binding induces conformational changes in the ATP-binding cassettes that increase their affinity for ATP.

3. ATP binds to the ATP-binding cassettes, changing their conformations so that the two domains interact strongly with one another. The close interaction of the ABCs reorients the transmembrane helices such that the substrate binding site is now facing outside the cell (Figure 13.8, far right).

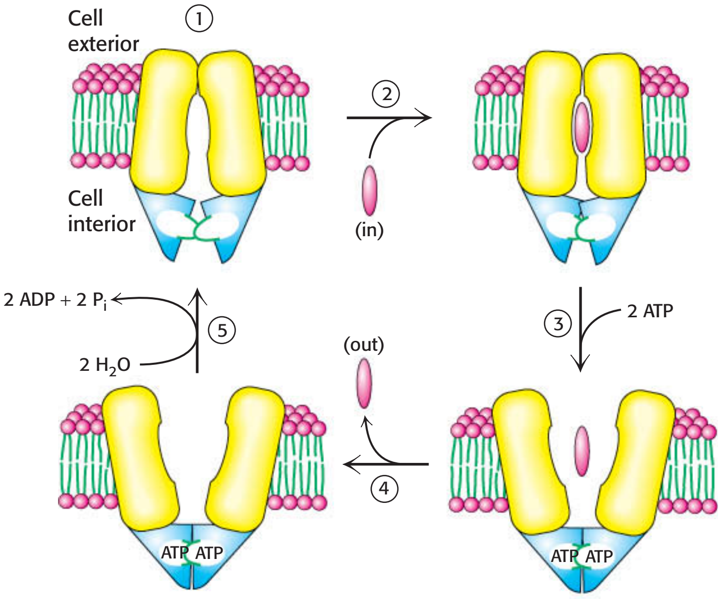

Cell exterior
①
Cell interior
②
(in)
③
2 ATP
④
(out)
⑤
2 ADP + 2 P_i
2 H_2O
ATP ATP
ATP ATP

FIGURE 13.9 ABC transporter mechanism. The mechanism includes the following steps: (1) opening of the channel toward the inside of the cell, (2) substrate binding and conformational changes in the ATP-binding cassettes, (3) ATP binding and opening of the channel to the opposite face of the membrane, (4) release of the substrate to the outside of the cell, and (5) ATP hydrolysis to reset the transporter to its initial state.

4. The outward facing conformation of the transporter has reduced affinity for the substrate, enabling the release of the substrate on the opposite face of the membrane.

5. The hydrolysis of ATP and the release of ADP and inorganic phosphate reset the transporter for another cycle.

Whereas eukaryotic ABC transporters generally act to export molecules from inside the cell, prokaryotic ABC transporters often act to import specific molecules from *outside* the cell. A specific binding protein acts in concert with the bacterial ABC transporter, delivering the substrate to the transporter and stimulating ATP hydrolysis inside the cell. These binding proteins are present in the periplasm, the compartment between the two membranes that surround some bacterial cells (Figure 12.35A).

Thus, ABC transporters use a substantially different mechanism from the P-type ATPases to couple the ATP hydrolysis reaction to conformational changes. Nonetheless, the net result is the same: the transporters are converted from one conformation capable of binding substrate from one side of the membrane to another that releases the substrate on the other side.

13.3 Lactose Permease Is an Archetype of Secondary Transporters That Use One Concentration Gradient to Power the Formation of Another

Carriers are proteins that transport ions or molecules across the membrane without hydrolysis of ATP. The mechanism of carriers involves both large conformational changes and the interaction of the protein with only a few molecules per transport cycle, limiting the maximum rate at which transport can occur. Although carriers cannot mediate primary active transport, owing to their inability to hydrolyze ATP, they can couple the thermodynamically unfavorable flow of one species of ion or molecule *up* a concentration gradient to the favorable flow of a different species *down* a concentration gradient, a process referred to as secondary active transport. Carriers that move ions or molecules "uphill" by this means are termed *secondary transporters* or *cotransporters*. These proteins can be classified as either *antiporters* or *symporters*. Antiporters couple the downhill flow of one species to the uphill flow of another in the *opposite direction* across the membrane; symporters use the flow of one species to drive the flow of a different species in the *same direction* across the membrane. *Uniporters,* another class of carriers, are able to transport a specific species in either direction governed only by concentrations of that species on either side of the membrane (Figure 13.10).

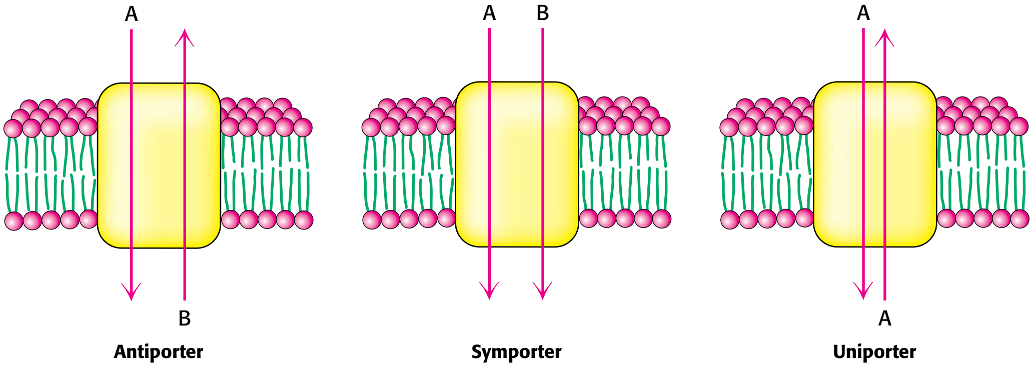


FIGURE 13.10 Antiporters, symporters, and uniporters. Secondary transporters can transport two substrates in opposite directions (antiporters), two substrates in the same direction (symporters), or one substrate in either direction (uniporter).

Secondary transporters are ancient molecular machines, common today in bacteria and archaea as well as in eukaryotes. For example, approximately 160 (of around 4000) proteins encoded by the *E. coli* genome are secondary transporters. Sequence comparison and hydropathy analysis suggest that members of the largest family have 12 transmembrane helices that appear to have arisen by duplication and fusion of a membrane protein containing 6 transmembrane helices. Included in this family is the *lactose permease* of *E. coli*. This symporter uses the H^+ gradient across the *E. coli* membrane (outside has higher H^+ concentration) generated by the oxidation of fuel molecules to drive the uptake of lactose and other sugars against a concentration gradient. This transporter has been extensively studied for many decades and is a useful archetype for this family.

The structure of lactose permease has been determined (Figure 13.11). As expected from the sequence analysis, the protein consists of two halves, each of which comprises six membrane-spanning α helices. The two halves are well separated and are joined by a single stretch of polypeptide. In this structure, a sugar molecule lies in a pocket in the center of the protein and is accessible from a path that leads from the interior of the cell. On the basis of this structure and a wide range of other experiments, a mechanism for symporter action has been developed. This mechanism (Figure 13.12) has many features similar to those for P-type ATPases and ABC transporters:

1. The cycle begins with the two halves oriented so that the opening to the binding pocket faces outside the cell, in a conformation different from that observed in the structures solved to date. A proton from outside the cell binds to a residue in the permease, quite possibly Glu 269.
2. In the protonated form, the permease binds lactose from outside the cell.
3. The structure everts to the form observed in the crystal structure (Figure 13.11).
4. The permease releases lactose to the inside of the cell.
5. The permease releases a proton to the inside of the cell.
6. The permease everts to complete the cycle.

The site of protonation likely changes in the course of this cycle.

It is believed that this eversion mechanism applies to all classes of secondary transporters, which resemble the lactose permease in overall architecture.

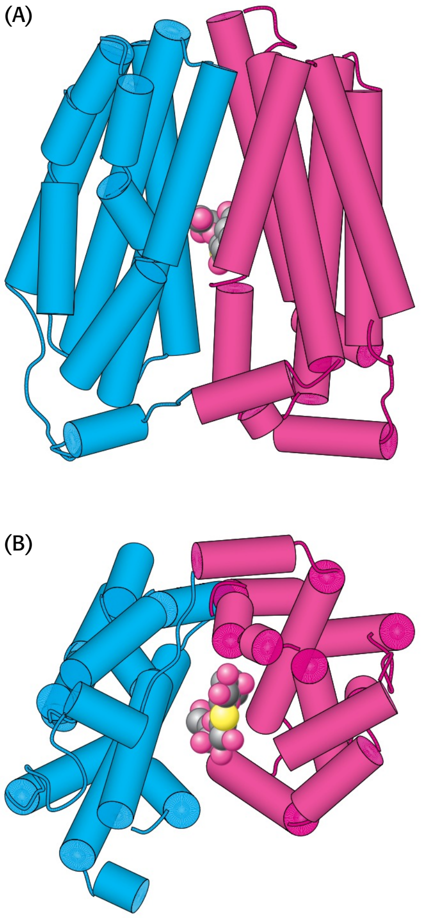


FIGURE 13.11 Structure of lactose permease with a bound lactose analog. The amino-terminal half of the protein is shown in blue and the carboxyl-terminal half in red. (A) Side view. (B) Bottom view (from inside the cell). *Notice* that the structure consists of two halves that surround the sugar and are linked to one another by only a single stretch of polypeptide. [Drawn from 1PV7.pdb.]

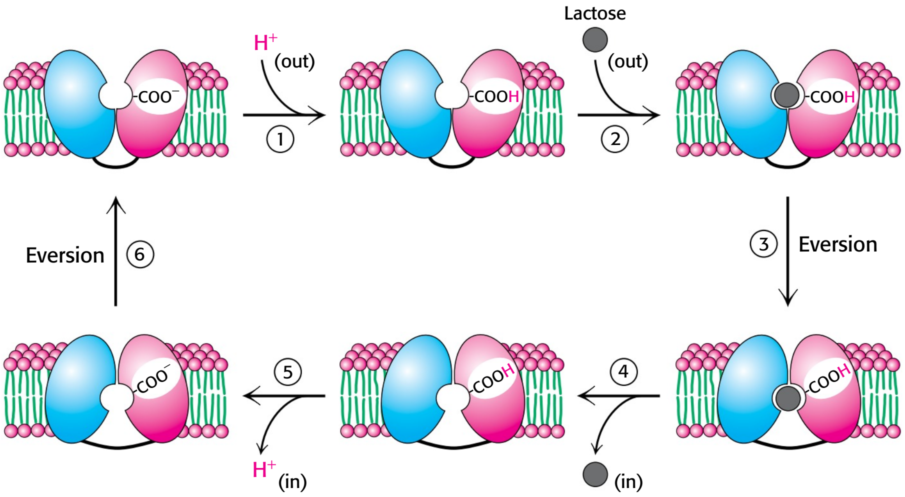


FIGURE 13.12 Lactose permease mechanism. The mechanism begins with the permease open to the outside of the cell (upper left). The permease binds a proton from the outside of the cell (1) and then binds its substrate (2). The permease everts (3) and then releases its substrate (4) and a proton (5) to the inside of the cell. It then everts (6) to complete the cycle.

13.4 Specific Channels Can Rapidly Transport Ions Across Membranes

Pumps and carriers can move ions across the membrane at rates approaching several thousand ions per second. Other membrane proteins, the passive-transport systems called *ion channels*, are capable of ion-transport rates that are more than 1000 times as fast. These rates of transport through ion channels are close to rates expected for ions diffusing freely through aqueous solution. Yet ion channels are not simply tubes that span membranes through which ions can rapidly flow. Instead, they are highly sophisticated molecular machines that respond to chemical and physical changes in their environments and undergo precisely timed conformational changes.

Action potentials are mediated by transient changes in Na^+ and K^+ permeability

One of the most important manifestations of ion-channel action is the nerve impulse, which is the fundamental means of communication in the nervous system. A *nerve impulse* is an electrical signal produced by the flow of ions across the plasma membrane of a neuron. The interior of a neuron, like that of most other cells, contains a high concentration of K^+ and a low concentration of Na^+. These ionic gradients are generated by the Na^+–K^+ ATPase. The cell membrane has an electrical potential determined by the ratio of the internal to the external concentration of ions. In the resting state, the membrane potential is typically −60 mV. A nerve impulse, or *action potential*, is generated when the membrane potential is depolarized beyond a critical threshold value (e.g., from −60 to −40 mV). The membrane potential becomes positive within about a millisecond and attains a value of about +30 mV before turning negative again (repolarization). This amplified depolarization is propagated along the nerve terminal (Figure 13.13).

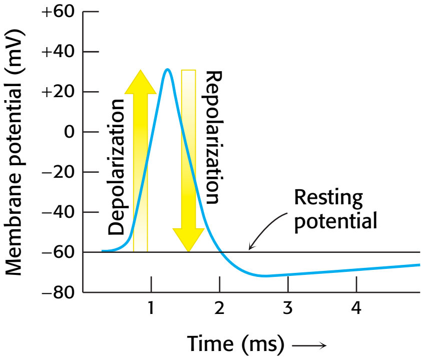


FIGURE 13.13 Action potential. Signals are sent along neurons by the transient depolarization and repolarization of the membrane.

Ingenious experiments carried out by Alan Hodgkin and Andrew Huxley revealed that action potentials arise from large, transient changes in the permeability of the axon membrane to Na^+ and K^+ ions. Depolarization of the membrane beyond the threshold level leads to an increase in permeability to Na^+. Sodium ions begin to flow into the cell because of the large electrochemical gradient across the plasma membrane. The entry of Na^+ further depolarizes the membrane, leading to a further increase in Na^+ permeability. This positive feedback yields the very rapid and large change in membrane potential described above and shown in Figure 13.13.

The membrane spontaneously becomes less permeable to Na^+ and more permeable to K^+. Consequently, K^+ flows outward, and so the membrane potential returns to a negative value. The resting level of −60 mV is restored in a few milliseconds as the K^+ conductance decreases to the value characteristic of the unstimulated state. The wave of depolarization followed by repolarization moves rapidly along a nerve cell. The propagation of these waves allows a touch at the tip of your toe to be detected in your brain in a few milliseconds.

This model for the action potential postulated the existence of ion channels specific for Na^+ and K^+. These channels must open in response to changes in membrane potential and then close after having remained open for a brief period of time. This bold hypothesis predicted the existence of molecules with a well-defined set of properties long before tools existed for their direct detection and characterization.

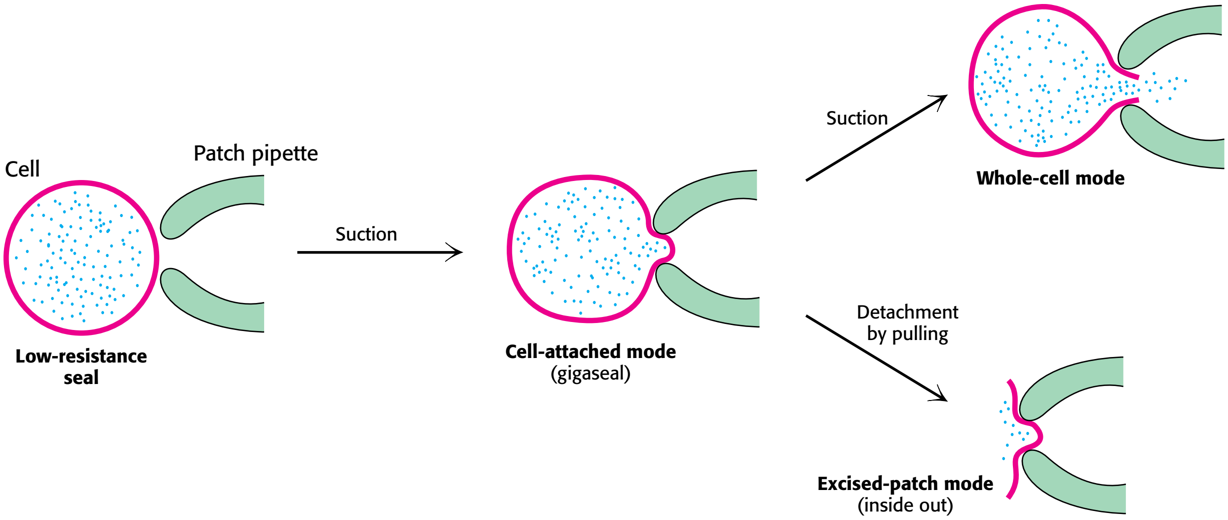


FIGURE 13.14 Patch-clamp modes. The patch-clamp technique for monitoring channel activity is highly versatile. A high-resistance seal (gigaseal) is formed between the pipette and a small patch of plasma membrane. This configuration is called *cell-attached mode*. The breaking of the membrane patch by increased suction produces a low-resistance pathway between the pipette and the interior of the cell. The activity of the channels in the entire plasma membrane can be monitored in this *whole-cell mode*. To prepare a membrane in the *excised-patch mode*, the pipette is pulled away from the cell. A piece of plasma membrane with its cytoplasmic side now facing the medium is monitored by the patch pipette.

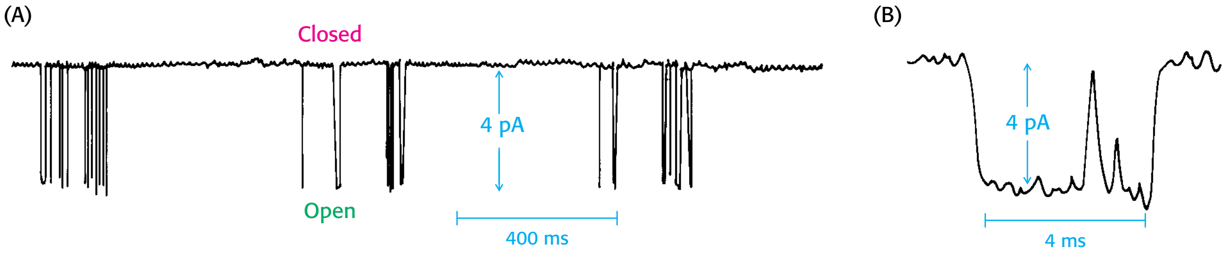
Patch-clamp conductance measurements reveal the activities of single channels

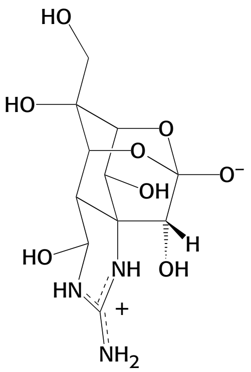
Direct evidence for the existence of these channels was provided by the *patch-clamp technique,* which was introduced by Erwin Neher and Bert Sakmann in 1976. This powerful technique enables the measurement of the ion conductance through a small patch of cell membrane. In this technique, a clean glass pipette with a tip diameter of about 1 μm is pressed against an intact cell to form a seal (Figure 13.14). Slight suction leads to the formation of a very tight seal so that the resistance between the inside of the pipette and the bathing solution is many gigaohms (1 gigaohm is equal to 10^9 ohms). Thus, a gigaohm seal (called a *gigaseal*) ensures that an electric current flowing through the pipette is identical with the current flowing through the membrane covered by the pipette. The gigaseal makes possible high-resolution current measurements while a known voltage is applied across the membrane. Remarkably, the flow of ions through a single channel and transitions between the open and the closed states of a channel can be monitored with a time resolution of microseconds (Figure 13.15). Furthermore, the activity of a channel in its native membrane environment, even in an intact cell, can be directly observed. Patch-clamp methods provided one of the first views of single biomolecules in action. Subsequently, other methods for observing single molecules were invented, opening new vistas on biochemistry at its most fundamental level.

The structure of a potassium ion channel is an archetype for many ion-channel structures

With the existence of ion channels firmly established by patch-clamp methods, scientists sought to identify the molecules that form ion channels. The Na^+ channel was first purified from the electric organ of electric eel,

FIGURE 13.15 Observing single channels. (A) The results of a patch-clamp experiment revealing the small amount of current, measured in picoamperes (pA, 10^{-12} amperes) passing through a single ion channel. The downward spikes indicate transitions between closed and open states. (B) Closer inspection of one of the spikes in (A) reveals the length of time the channel is in the open state.





Tetrodotoxin

which is a rich source of the protein forming this channel. The channel was purified on the basis of its ability to bind tetrodotoxin, a neurotoxin from the puffer fish that binds to Na^+ channels very tightly ($K_\text{i} \approx 1$ nM). The lethal dose of this poison for an adult human being is about 10 ng.

The isolated Na^+ channel is a single 260-kDa chain. Cloning and sequencing of cDNAs encoding Na^+ channels revealed that the channel contains four internal repeats, each having a similar amino acid sequence, suggesting that gene duplication and divergence have produced the gene for this channel. Hydrophobicity profiles indicate that each repeat contains five hydrophobic segments (S1, S2, S3, S5, and S6). Each repeat also contains a highly positively charged S4 segment; positively charged arginine or lysine residues are present at nearly every third residue. It was proposed that segments S1 through S6 are membrane-spanning α helices, while the positively charged residues in S4 act as the voltage sensors of the channel.

The purification of K^+ channels proved to be much more difficult because of their low abundance and the lack of known high-affinity ligands comparable to tetrodotoxin. The breakthrough came in studies of mutant fruit flies that shake violently when anesthetized with ether. The mapping and cloning of the gene, termed *shaker,* responsible for this defect revealed the amino acid sequence encoded by a K^+-channel gene. The *shaker* gene encodes a 70-kDa protein that contains sequences corresponding to segments S1 through S6 in one of the repeated units of the Na^+ channel. Thus, a K^+-channel subunit is homologous to one of the repeated units of Na^+ channels. Consistent with this homology, four Shaker polypeptides come together to form a functional channel. More recently, bacterial K^+ channels were discovered that contain only the two membrane-spanning regions corresponding to segments S5 and S6. This and other information suggested that S5 and S6, including the region between them, form the actual pore in the K^+ channel. Segments S1 through S4 contain the apparatus that opens the pore. The sequence relations between these ion channels are summarized in Figure 13.16.

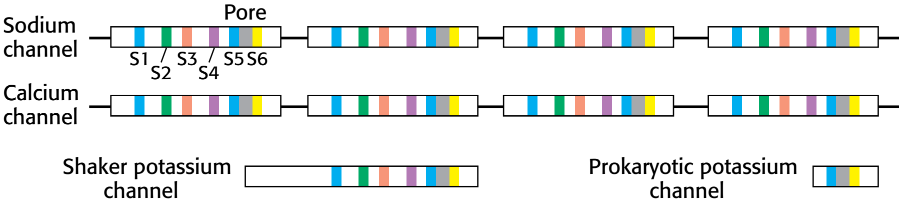


FIGURE 13.16 Sequence relations of ion channels. Like colors indicate structurally similar regions of the sodium, calcium, and potassium channels. Each of these channels exhibits approximate fourfold symmetry, either within one chain (sodium, calcium channels) or by forming tetramers (potassium channels).

In 1998, Roderick MacKinnon and coworkers determined the structure of a K^+ channel from the bacterium *Streptomyces lividans* by x-ray crystallography. This channel contains only the pore-forming segments S5 and S6. As expected, the K^+ channel is a tetramer of identical subunits, each of which includes two membrane-spanning α helices (Figure 13.17). The four subunits come together to form a pore in the shape of a cone that runs through the center of the structure.

The structure of the potassium ion channel reveals the basis of ion specificity

The structure presented in Figure 13.17 probably represents the K^+ channel in a closed form. Nonetheless, it suggests how the channel is able to exclude all but K^+ ions. Beginning from the inside of the cell, the pore starts with a diameter of approximately 10 Å and then constricts to a smaller cavity with a

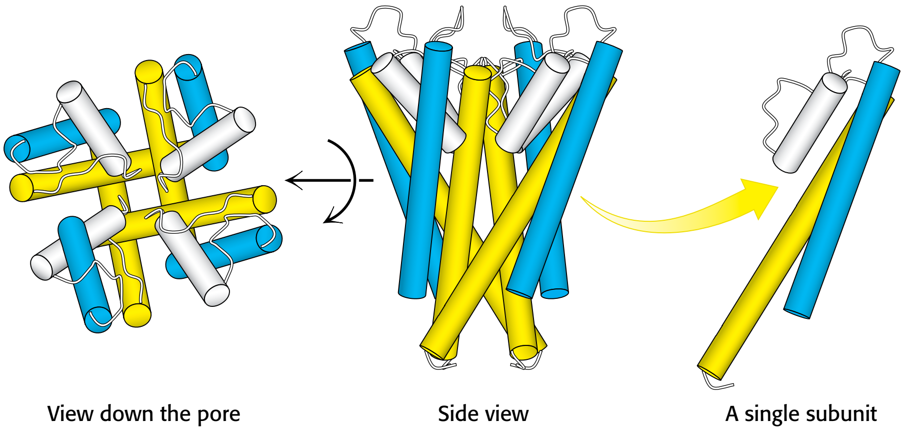


FIGURE 13.17 Structure of the potassium ion channel. The K^+ channel, composed of four identical subunits, is cone shaped, with the larger opening facing the inside of the cell (center). A view down the pore, looking toward the outside of the cell, shows the relations of the individual subunits (left). One of the four identical subunits of the pore is illustrated at the right, with the pore-forming region shown in gray. [Drawn from 1K4C.pdb.]

diameter of 8 Å. Both the opening to the outside and the central cavity of the pore are filled with water, and a K^+ ion can fit in the pore without losing its shell of bound water molecules. Approximately two-thirds of the way through the membrane, the pore becomes more constricted (3-Å diameter). At that point, any K^+ ions must give up their water molecules and interact directly with groups from the protein. The channel structure effectively reduces the thickness of the membrane from 34 Å to 12 Å by allowing the solvated ions to penetrate into the membrane before the ions must directly interact with the channel (Figure 13.18).

For K^+ ions to relinquish their water molecules, other polar interactions must replace those with water. The restricted part of the pore is built from residues contributed by the two transmembrane α helices. In particular, a five-amino-acid stretch within this region functions as the *selectivity filter* that determines the preference for K^+ over other ions (Figure 13.19). The stretch has the sequence Thr-Val-Gly-Tyr-Gly (TVGYG), and is nearly completely conserved in all K^+ channels. The region of the strand containing the conserved sequence lies in an extended conformation and is oriented such that the peptide carbonyl groups are directed into the channel, in good position to interact with the potassium ions.

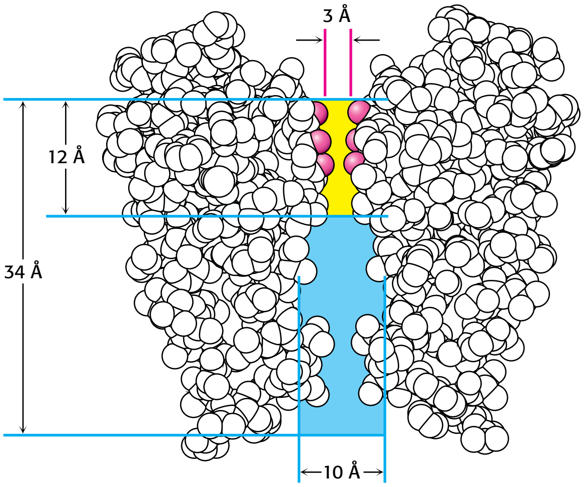


FIGURE 13.18 Path through a channel. A potassium ion entering the K^+ channel can pass a distance of 22 Å into the membrane while remaining solvated with water (blue). At this point, the pore diameter narrows to 3 Å (yellow), and potassium ions must shed their water and interact with carbonyl groups (red) of the pore amino acids.

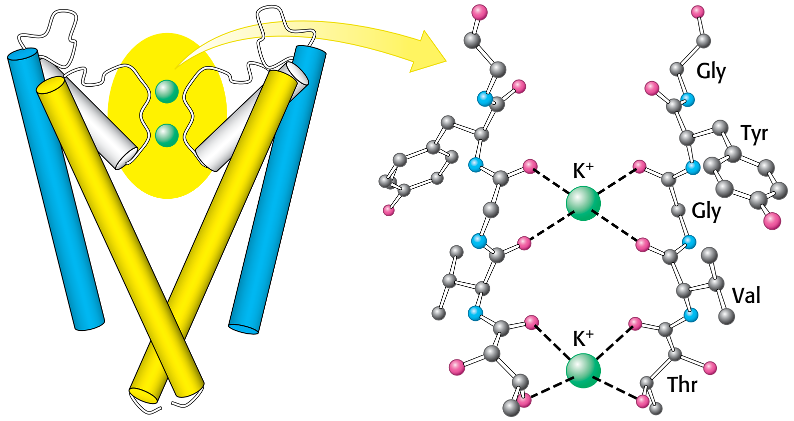


FIGURE 13.19 Selectivity filter of the potassium ion channel. Potassium ions interact with the carbonyl groups of the TVGYG sequence of the selectivity filter, located at the 3-Å-diameter pore of the K^+ channel. Only two of the four channel subunits are shown.

TABLE 13.1 Properties of alkali cations

Ion	Ionic radius (Å)	Hydration free energy in kJ mol^{-1} ($kcal\ mol^{-1}$)
Li^+	0.60	−410 (−98)
Na^+	0.95	−301 (−72)
K^+	1.33	−230 (−55)
Rb^+	1.48	−213 (−51)
Cs^+	1.69	−197 (−47)

Potassium ion channels are 100-fold more permeable to K^+ than to Na^+. How is this high degree of selectivity achieved? Ions having a radius larger than 1.5 Å cannot pass into the narrow diameter (3 Å) of the selectivity filter of the K^+ channel. However, a bare Na^+ is small enough (Table 13.1) to pass through the pore. Indeed, the ionic radius of Na^+ is substantially smaller than that of K^+. How then is Na^+ rejected?

The key point is that the free-energy costs of dehydrating these ions are considerable [Na^+, 301 kJ mol^{-1} (72 kcal mol^{-1}), and K^+, 230 kJ mol^{-1} (55 kcal mol^{-1})]. *The channel pays the cost of dehydrating K^+ by providing optimal compensating interactions with the carbonyl oxygen atoms lining the selectivity filter.* Careful studies of the potassium channel, enabled by the determination of its three-dimensional structure, have revealed that the interior of the pore is a highly dynamic, fluid environment. The favorable interactions between the carbonyl oxygen atoms, which carry a partial negative charge, with the cation are balanced by the repulsion of these oxygen atoms from one another. For this channel, the ideal balance is achieved with K^+, but not with Na^+ (Figure 13.20). Hence, sodium ions are rejected because the higher energetic cost of dehydrating them would not be recovered.

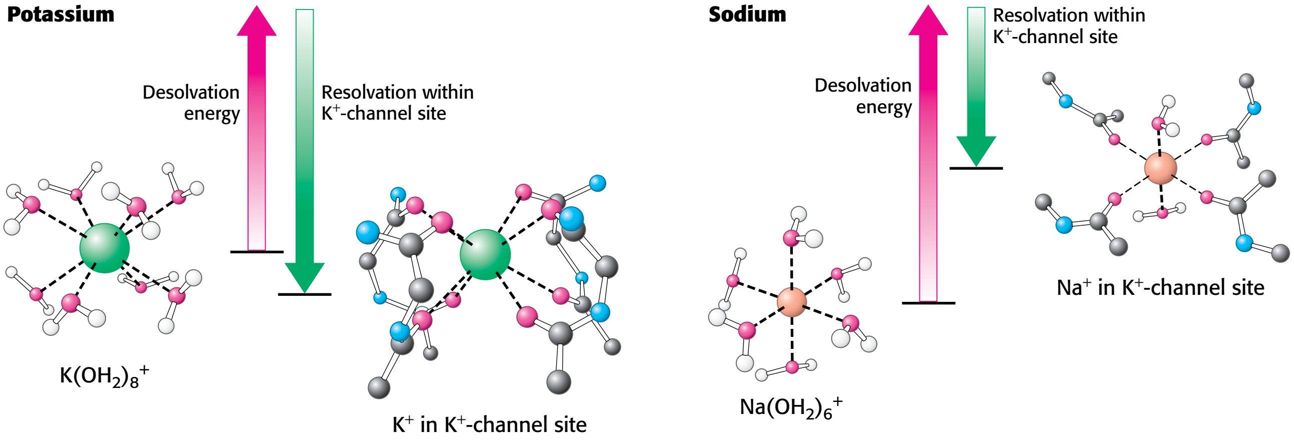

Potassium
Desolvation energy
Resolution within K^+-channel site
$K(OH_2)_8{}^+$
K^+ in K^+-channel site
Sodium
Desolvation energy
Resolution within K^+-channel site
$Na(OH_2)_6{}^+$
Na^+ in K^+-channel site

FIGURE 13.20 Energetic basis of ion selectivity. The energy cost of dehydrating a potassium ion is compensated by favorable interactions with the selectivity filter. Because a sodium ion is too small to interact favorably with the selectivity filter, the free energy of desolvation cannot be compensated and the sodium ion does not pass through the channel.

The K^+ channel structure enables a clearer understanding of the structure and function of Na^+ and Ca^{2+} channels because of their homology to K^+ channels. Sequence comparisons and the results of mutagenesis experiments have implicated the region between segments S5 and S6 in ion selectivity in the Ca^{2+} channel. In Ca^{2+} channels, one glutamate residue of this region in each of the four repeated units plays a major role in determining ion selectivity. Residues in the positions corresponding to the glutamate residues in Ca^{2+} channels are major components of the selectivity filter of the Na^+ channel. These residues—aspartate, glutamate, lysine, and alanine—are located in each of the internal repeats of the Na^+ channel, forming a region termed the DEKA locus. Thus, the potential fourfold symmetry of the channel is clearly broken in this region, which explains why Na^+ channels consist of a single large polypeptide chain rather than a noncovalent assembly of four identical subunits. The preference of the Na^+ channel for Na^+ over K^+ depends on ionic radius; the diameter of the pore determined by these residues and others is sufficiently restricted that small ions such as Na^+ and Li^+ can pass through the channel, but larger ions such as K^+ are significantly hindered.

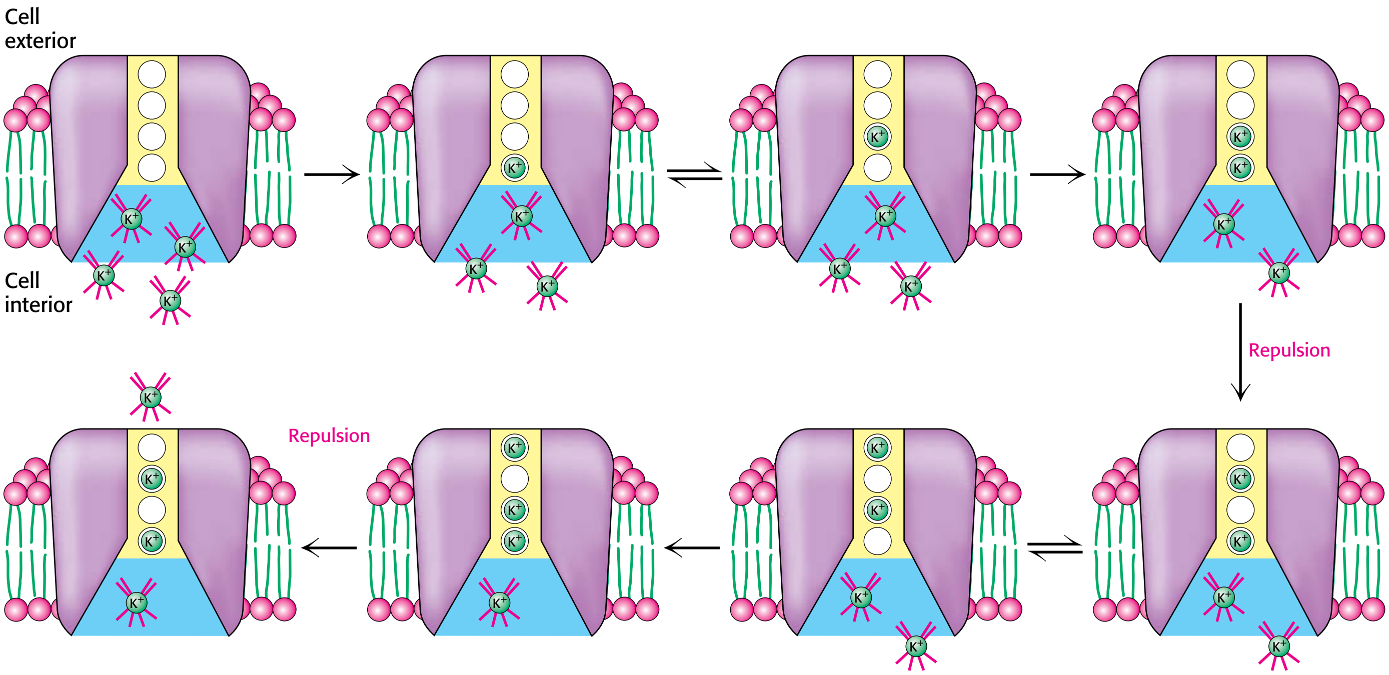


FIGURE 13.21 Model for K$^+$-channel ion transport. The selectivity filter has four binding sites. Hydrated potassium ions can enter these sites, one at a time, losing their hydration shells. When two ions occupy adjacent sites, electrostatic repulsion forces them apart. Thus, as ions enter the channel from one side, other ions are pushed out the other side.

The structure of the potassium ion channel explains its rapid rate of transport

The tight binding sites required for ion selectivity should slow the progress of ions through a channel, yet ion channels achieve rapid rates of ion transport. How is this paradox resolved? A structural analysis of the K$^+$ channel at high resolution provides an appealing explanation. Four K$^+$-binding sites crucial for rapid ion flow are present in the constricted region of the K$^+$ channel. Consider the process of ion conductance starting from inside the cell (Figure 13.21). A hydrated potassium ion proceeds into the channel and through the relatively unrestricted part of the channel. The ion then gives up its coordinated water molecules and binds to a site within the selectivity-filter region. The ion can move between the four sites within the selectivity filter because they have similar ion affinities. As each subsequent potassium ion moves into the selectivity filter, its positive charge will repel the potassium ion at the nearest site, causing it to shift to a site farther up the channel and in turn push upward any potassium ion already bound to a site farther up. Thus, each ion that binds anew favors the release of an ion from the other side of the channel. This multiple-binding-site mechanism solves the paradox of high ion selectivity and rapid flow.

Voltage gating requires substantial conformational changes in specific ion-channel domains

Some Na$^+$ and K$^+$ channels are gated by membrane potential; that is, they change conformation to a highly conducting form in response to changes in voltage across the membrane. As already noted, these *voltage-gated channels* include segments S1 through S4 in addition to the pore itself formed by S5 and S6. The structure of a voltage-gated K$^+$ channel from *Aeropyrum pernix* has been determined by x-ray crystallography (Figure 13.22). The segments S1 through S4 form domains, termed "paddles," that extend from the core of the channel. These paddles include the segment S4, the voltage sensor itself. Segment S4 forms an α helix lined with positively charged residues. In contrast with expectations, segments S1 through S4 are not enclosed within the protein but, instead, are positioned to lie in the membrane itself.

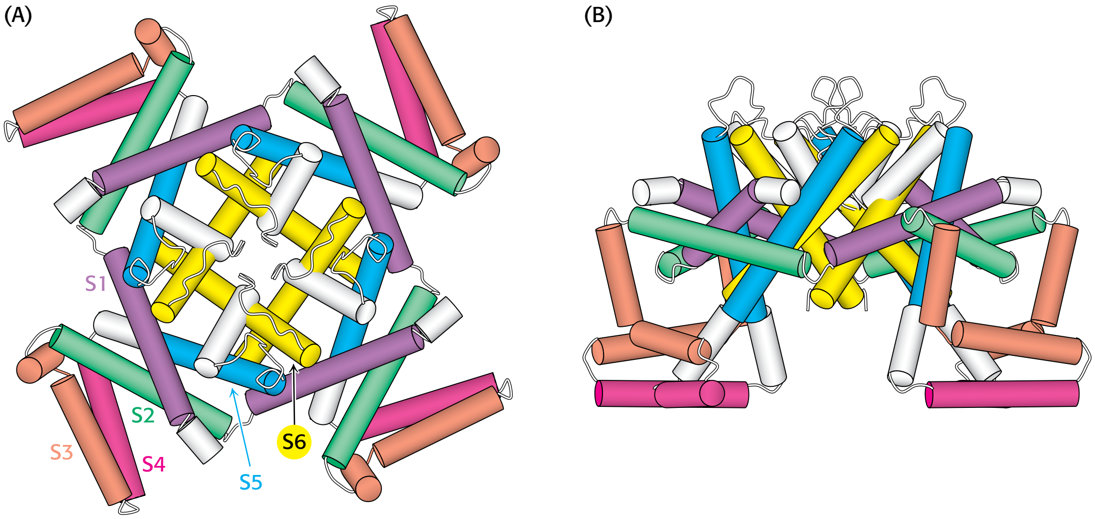
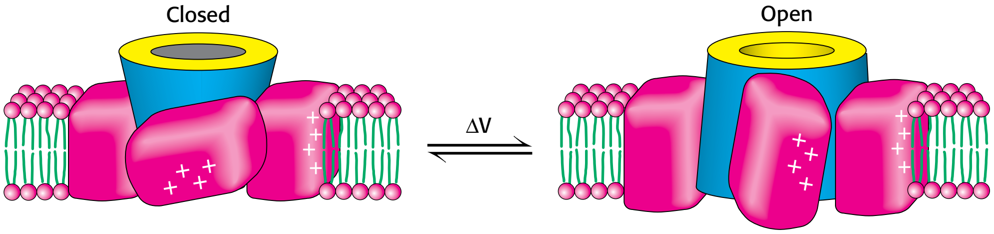


FIGURE 13.22 Structure of a voltage-gated potassium channel. (A) A view looking down through the pore. (B) A side view. *Notice* that the positively charged S4 region (red) lies on the outside of the structure at the bottom of the pore. [Drawn from 1ORQ.pdb.]

A model for voltage gating has been proposed by Roderick MacKinnon and coworkers on the basis of this structure and a range of other experiments (Figure 13.23). In the closed state, the paddles lie in a "down" position. On membrane depolarization, the cytoplasmic side of the membrane becomes more positively charged, and the paddles are pulled through the membrane into an "up" position. In this position, they pull the four sides of the base on the pore apart, increasing access to the selectivity filter and opening the channel.

FIGURE 13.23 A model for voltage gating of ion channels. The voltage-sensing paddles lie in the "down" position below the closed channel (left). Membrane depolarization pulls these paddles through the membrane. The motion pulls the base of the channel apart, opening the channel (right).



A channel can be inactivated by occlusion of the pore: the ball-and-chain model

The K^+ channel and the Na^+ channel undergo inactivation within milliseconds of opening (Figure 13.24). A first clue to the mechanism of inactivation came from exposing the cytoplasmic side of either channel to trypsin; cleavage by trypsin produced trimmed channels that stayed persistently open after depolarization. Furthermore, a mutant Shaker channel lacking 42 amino acids near the amino terminus opened in response to depolarization but did not inactivate. Remarkably, inactivation was restored by adding a synthetic peptide corresponding to the first 20 residues of the native channel.

These experiments strongly support the *ball-and-chain model* for channel inactivation that had been proposed years earlier (Figure 13.25). According to this model, the first 20 residues of the K^+ channel form a cytoplasmic unit (the *ball*) that is attached to a flexible segment of the polypeptide (the *chain*). When the channel is closed, the ball rotates freely in

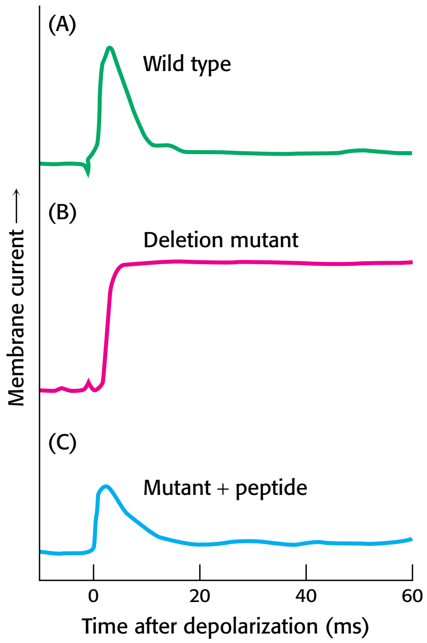
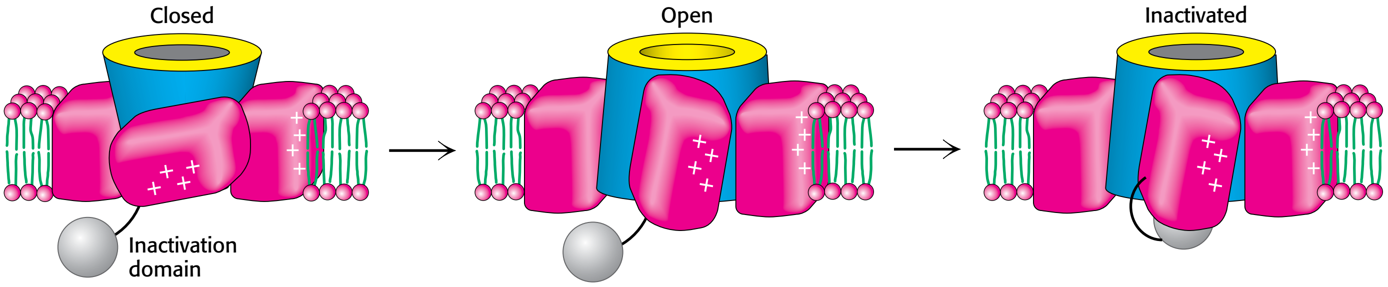


FIGURE 13.24 Inactivation of the potassium ion channel. The amino-terminal region of the K^+ chain is critical for inactivation. (A) The wild-type Shaker K^+ channel displays rapid inactivation after opening. (B) A mutant channel lacking residues 6 through 46 does not inactivate. (C) Inactivation can be restored by adding a peptide consisting of residues 1 through 20 at a concentration of 100 μM. [Data from W. N. Zagotta, T. Hoshi, and R. W. Aldrich, *Science* 250:568–571, 1990.]

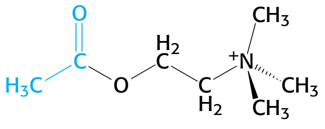
the aqueous solution. When the channel opens, the ball quickly finds a complementary site in the open pore and occludes it. Hence, the channel opens for only a brief interval before it undergoes inactivation by occlusion. Shortening the chain speeds inactivation because the ball finds its target more quickly. Conversely, lengthening the chain slows inactivation. Thus, the duration of the open state can be controlled by the length and flexibility of the tether. In some senses, the "ball" domains, which include substantial regions of positive charge, can be thought of as large, tethered cations that are pulled into the open channel but get stuck and block further ion conductance.

FIGURE 13.25 Ball-and-chain model for channel inactivation. The inactivation domain, or "ball" (gray), is tethered to the channel by a flexible "chain." In the closed state, the ball is located in the cytoplasm. Depolarization opens the channel and creates a binding site for the positively charged ball in the mouth of the pore. Movement of the ball into this site inactivates the channel by occluding it.



The acetylcholine receptor is an archetype for ligand-gated ion channels

Nerve impulses are communicated across synapses by small, diffusible molecules called *neurotransmitters*. One neurotransmitter is *acetylcholine*. The presynaptic membrane of a synapse is separated from the postsynaptic membrane by a gap of about 50 nm called the *synaptic cleft*. The arrival of a nerve impulse at the end of an axon leads to the synchronous export of the contents of some 300 membrane-bound compartments, or vesicles, of acetylcholine into the cleft (Figure 13.26). The binding of acetylcholine to the postsynaptic membrane markedly changes its ionic permeability, triggering an action potential. Acetylcholine opens a single kind of cation channel, called the *acetylcholine receptor,* which is almost equally permeable to Na^+ and to K^+.



Acetylcholine

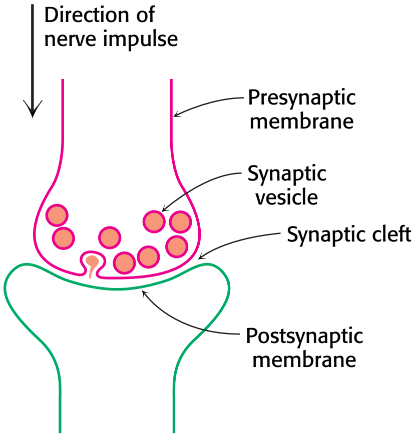


FIGURE 13.26 Schematic representation of a synapse.

The torpedo (*Torpedo marmorata*, also known as the electric ray) has an electric organ, rich in acetylcholine receptors, that can deliver a shock of as much as 200 V for approximately 1 s. [Yves Gladu/Jacana/Photo Researchers.]

The acetylcholine receptor is the best-understood *ligand-gated channel*. This type of channel is gated not by voltage but by the presence of specific ligands. The binding of acetylcholine to the channel is followed by its transient opening. The electric organ of *Torpedo marmorata*, an electric ray, is a choice source of acetylcholine receptors for study because its electroplaxes (voltage-generating cells) are very rich in postsynaptic membranes that respond to this neurotransmitter. The receptor is very densely packed in these membranes ($\sim 20{,}000\ \mu m^{-2}$). The acetylcholine receptor of the electric organ has been solubilized by adding a nonionic detergent to a post-synaptic membrane preparation and purified by affinity chromatography on a column bearing covalently attached cobratoxin, a small protein toxin from snakes that has a high affinity for acetylcholine receptors. With the use of techniques presented in Chapter 3, the 268-kDa receptor was identified as a pentamer of four kinds of membrane-spanning subunits—α_2, β, γ, and δ—arranged in the form of a ring that creates a pore through the membrane.

The cloning and sequencing of the cDNAs for the four kinds of subunits (50–58 kDa) showed that they have clearly similar sequences; the genes for the α, β, γ, and δ subunits arose by duplication and divergence of a common ancestral gene. Each subunit has a large extracellular domain, followed at the carboxyl end by four predominantly hydrophobic segments that span the bilayer membrane. Acetylcholine binds at the α−γ and α−δ interfaces. Electron microscopic studies of purified acetylcholine receptors demonstrated that the structure has approximate fivefold symmetry, in harmony with the similarity of its five constituent subunits (Figure 13.27).

What is the basis of channel opening? A comparison of the structures of the closed and open forms of the channel would be highly revealing, but such comparisons have been difficult to obtain. Cryoelectron micrographs indicate that the binding of acetylcholine to the extracellular domain causes a structural alteration that initiates rotations of the α-helical rods lining the membrane-spanning pore. The amino acid sequences of these helices point to the presence of alternating ridges of small polar or neutral residues (serine, threonine, glycine) and large nonpolar ones (isoleucine,

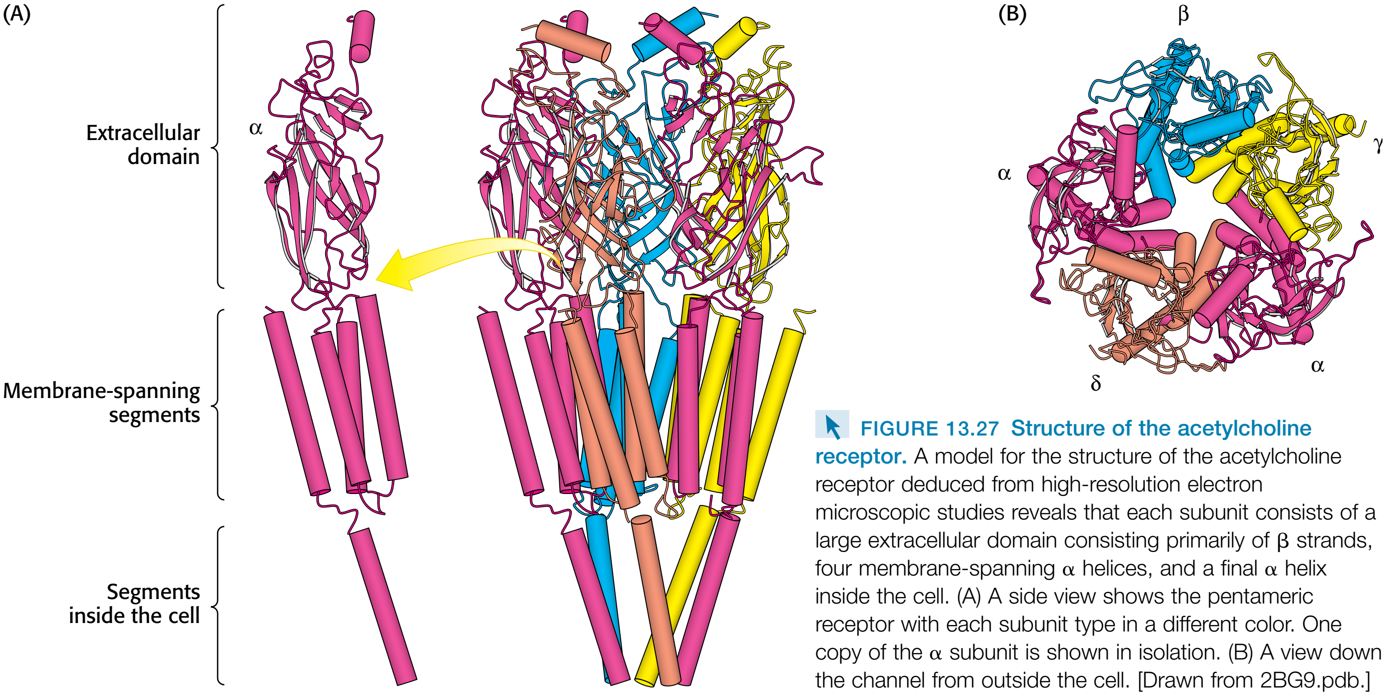


FIGURE 13.27 Structure of the acetylcholine receptor. A model for the structure of the acetylcholine receptor deduced from high-resolution electron microscopic studies reveals that each subunit consists of a large extracellular domain consisting primarily of β strands, four membrane-spanning α helices, and a final α helix inside the cell. (A) A side view shows the pentameric receptor with each subunit type in a different color. One copy of the α subunit is shown in isolation. (B) A view down the channel from outside the cell. [Drawn from 2BG9.pdb.]

leucine, phenylalanine). In the closed state, the large residues may occlude the channel by forming a tight hydrophobic ring (Figure 13.28). Indeed, each subunit has a bulky leucine residue at a critical position. The binding of acetylcholine could allosterically rotate the membrane-spanning helices so that the pore would be lined by small polar residues rather than by large hydrophobic ones. The wider, more polar pore would then be open to the passage of Na^+ and K^+ ions.

Action potentials integrate the activities of several ion channels working in concert

To see how ligand-gated and voltage-gated channels work together to generate a sophisticated physiological response, we now revisit the action potential introduced at the beginning of this section. First, we need to introduce the concept of *equilibrium potential*. Suppose that a membrane separates two solutions that contain different concentrations of some cation X^+, as well as an equivalent amount of anions to balance the charge in each solution (Figure 13.29). Let $[X^+]_{in}$ be the concentration of X^+ on one side of the membrane (corresponding to the inside of a cell) and $[X^+]_{out}$ be the concentration of X^+ on the other side (corresponding to the outside of a cell). Suppose that an ion channel opens that allows X^+ to move across the membrane. What will happen? It seems clear that X^+ will move through the channel from the side with the higher concentration to the side with the lower concentration. However, positive charges will start to accumulate on the side with the lower concentration, making it more difficult to move each additional positively charged ion. An equilibrium will be achieved when the driving force due to the concentration gradient is balanced by the electrostatic force resisting the motion of an additional charge. In these circumstances, the membrane potential is given by the *Nernst equation:*

$$V_{eq} = -(RT/zF)\ln([X]_{in}/[X]_{out})$$

where R is the gas constant and F is the Faraday constant (96.5 $kJ\,V^{-1}\,mol^{-1}$, or 23.1 $kcal\,V^{-1}\,mol^{-1}$) and z is the charge on the ion X (e.g., +1 for X^+).

The membrane potential at equilibrium is called the equilibrium potential for a given ion at a given concentration ratio across a membrane. For sodium with $[Na^+]_{in} = 14$ mM and $[Na^+]_{out} = 143$ mM, the equilibrium potential is +62 mV at 37°C. Similarly, for potassium with $[K^+]_{in} = 157$ mM and $[K^+]_{out} = 4$ mM, the equilibrium potential is −98 mV. In the absence of

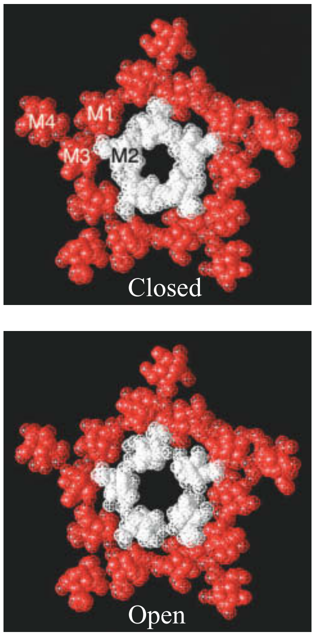


FIGURE 13.28 Opening the acetylcholine receptor. Cross sections from electron microscopic reconstructions of the acetylcholine receptor in (top) its closed form and (bottom) its open form. (The open form corresponds to the structure shown in Figure 13.27). The areas labeled M1, M2, M3, and M4 correspond to the four membrane-spanning α helices of one subunit. The cross section of the open channel was generated by treating the receptor with acetylcholine and freezing the sample within 20 ms. *Notice* that the hole in the center of the channel is substantially larger in the open structure. The enlargement of the hole is due to the rotation of the M2 helices by approximately 15 degrees along their long axes. [Courtesy of Nigel Unwin.]

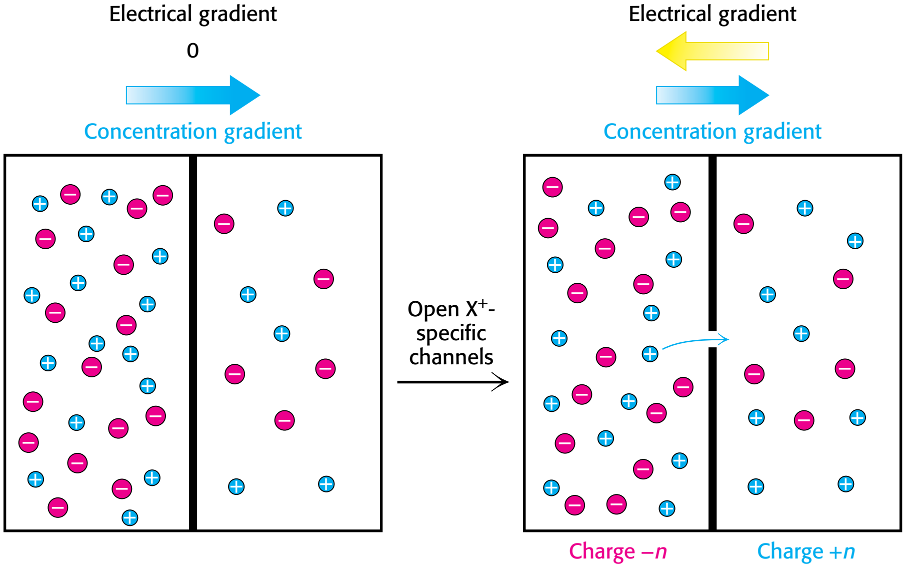


FIGURE 13.29 Equilibrium potential. The membrane potential reaches equilibrium when the driving force due to the concentration gradient is exactly balanced by the opposing force due to the repulsion of like charges.

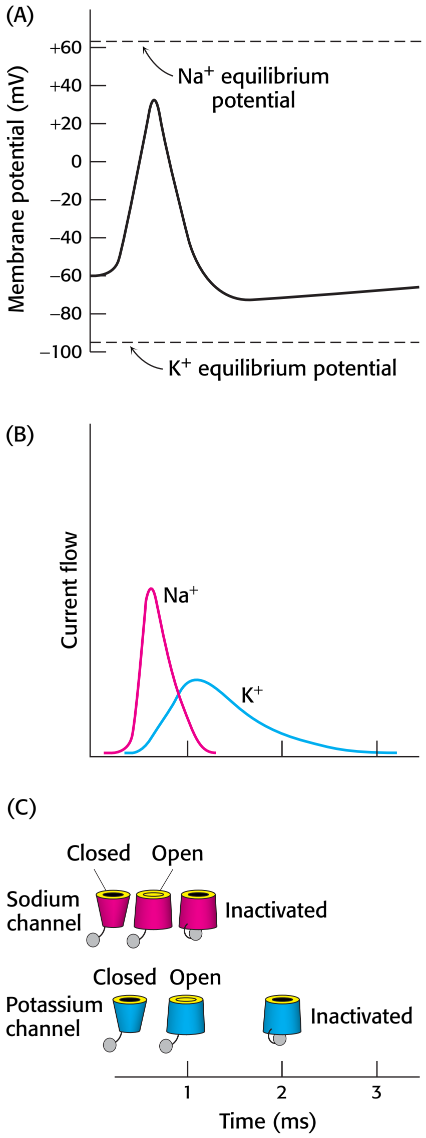


FIGURE 13.30 Action-potential mechanism. (A) On the initiation of an action potential, the membrane potential moves from the resting potential upward toward the Na^+ equilibrium potential and then downward toward the K^+ equilibrium potential. (B) The currents through the Na^+ and K^+ channels underlying the action potential. (C) The states of the Na^+ and K^+ channels during the action potential.

stimulation, the resting potential for a typical neuron is −60 mV. This value is close to the equilibrium potential for K^+ owing to the fact that a small number of K^+ channels are open.

We are now prepared to consider what happens in the generation of an action potential (Figure 13.30). Initially, a neurotransmitter such as acetylcholine is released into the synaptic cleft from a presynaptic membrane (Figure 13.26). The released acetylcholine binds to the acetylcholine receptor on the postsynaptic membrane, causing it to open within less than a millisecond. The acetylcholine receptor is a nonspecific cation channel. Sodium ions flow into the cell and potassium ions flow out of the cell. Without any further events, the membrane potential would move to a value corresponding to the average of the equilibrium potentials for Na^+ and K^+, approximately −20 mV. However, as the membrane potential approaches −40 mV, the voltage-sensing paddles of Na^+ channels are pulled into the membrane, opening the Na^+ channels. With these channels open, sodium ions flow rapidly into the cell and the membrane potential rises rapidly toward the Na^+ equilibrium potential (Figure 13.30B, red curve). The voltage-sensing paddles of K^+ channels also are pulled into the membrane by the changed membrane potential, but more slowly than Na^+ channel paddles. Nonetheless, after approximately 1 ms, many K^+ channels start to open. At the same time, inactivation "ball" domains plug the open Na^+ channels, decreasing the Na^+ current. The acetylcholine receptors that initiated these events are also inactivated on this time scale. With the Na^+ channels inactivated and only the K^+ channels open, the membrane potential drops rapidly toward the K^+ equilibrium potential (Figure 13.30B, blue curve). The open K^+ channels are susceptible to inactivation by their "ball" domains, and these K^+ currents, too, are blocked. With the membrane potential returned to close to its initial value, the inactivation domains are released and the channels return to their original closed states. These events propagate along the neuron as the depolarization of the membrane opens channels in nearby patches of membrane.

How much current actually flows across the membrane over the course of an action potential? This question can be addressed from two complementary directions. First, a typical nerve cell contains 100 Na^+ channels per square micrometer. At a membrane potential of +20 mV, each channel conducts 10^7 ions per second. Thus, in a period of 1 millisecond, approximately 10^5 ions flow through each square micrometer of membrane surface. Assuming a cell volume of 10^4 μm^3 and a surface area of 10^4 μm^2, this rate of ion flow corresponds to an increase in the Na^+ concentration of less than 1%. How can this be? A robust action potential is generated because the membrane potential is very sensitive to even a slight change in the distribution of charge. This sensitivity makes the action potential a very efficient means of signaling over long distances and with rapid repetition rates.

Disruption of ion channels by mutations or chemicals can be potentially life-threatening

The generation of an action potential requires the precise coordination of gating events of a collection of ion channels. Perturbation of this timing can have devastating effects. For example, the rhythmic generation of action potentials by the heart is absolutely essential to maintain delivery of oxygenated blood to the peripheral tissues. *Long QT syndrome* (LQTS) is a genetic disorder in which the recovery of the action potential from its peak potential to the resting equilibrium potential is delayed. The term "QT" refers to a specific feature of the cardiac electrical activity pattern as measured by electrocardiography. LQTS can lead to brief losses of consciousness (syncope), disruption of normal cardiac rhythm (arrhythmia), and sudden death. The most common mutations identified in LQTS patients inactivate K^+ channels or prevent the proper trafficking of these channels to the plasma membrane. The resulting loss in potassium permeability slows the repolarization of the membrane and delays the induction of the subsequent cardiac contraction, rendering the cardiac tissue susceptible to arrhythmias.

Prolongation of the cardiac action potential in this manner can also be induced by a number of therapeutic drugs. In particular, the K^+ channel hERG (for human *ether-a-go-go*-related gene, named for its ortholog in *Drosophila melanogaster*) is highly susceptible to interactions with certain drugs. The hydrophobic regions of these drugs can block hERG by binding to two nonconserved aromatic residues on the internal surface of the channel cavity. In addition, this cavity is predicted to be wider than other K^+ channels because of the absence of a conserved Pro-X-Pro motif within the S6 hydrophobic segment. Inhibition of hERG by these drugs can lead to an increased risk of cardiac arrhythmias and sudden death. Accordingly, a number of these agents, such as the antihistamine terfenadine, have been withdrawn from the market. Screening for the inhibition of hERG is now a critical safety hurdle for the pharmaceutical advancement of a molecule to an approved drug.

13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells

The ion channels that we have considered thus far have narrow pores and are moderately to highly selective in the ions that they allow to pass through them. They are closed in the resting state and have short lifetimes in the open state, typically a millisecond, that enable them to transmit frequent neural signals. We turn now to a channel with a very different role. *Gap junctions*, also known as *cell-to-cell channels*, serve as passageways between the interiors of contiguous cells. Gap junctions are clustered in discrete regions of the plasma membranes of apposed cells. Electron micrographs of sheets of gap junctions show them tightly packed in a regular hexagonal array (Figure 13.31). An approximately 20-Å central hole, the lumen of the channel, is prominent in each gap junction. These channels span the intervening space, or gap, between apposed cells (hence, the name "gap junction"). The width of the gap between the cytoplasms of the two cells is about 35 Å.

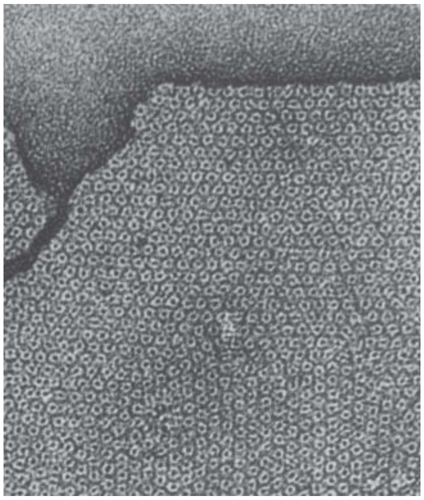


FIGURE 13.31 Gap junctions. This electron micrograph shows a sheet of isolated gap junctions. The cylindrical connexons form a hexagonal lattice having a unit-cell length of 85 Å. The densely stained central hole has a diameter of about 20 Å. [Don W. Fawcett/Photo Researchers.]

Small hydrophilic molecules as well as ions can pass through gap junctions. The pore size of the junctions was determined by microinjecting a series of fluorescent molecules into cells and observing their passage into adjoining cells. All polar molecules with a mass of less than about 1 kDa can readily pass through these cell-to-cell channels. Thus, *inorganic ions and most metabolites (e.g., sugars, amino acids, and nucleotides) can flow between the interiors of cells joined by gap junctions*. In contrast, proteins, nucleic acids, and polysaccharides are too large to traverse these channels. *Gap junctions are important for intercellular communication*. Cells in some excitable tissues, such as heart muscle, are coupled by the rapid flow of ions through these junctions, which ensures a rapid and synchronous response to stimuli. Gap junctions are also essential for the nourishment of cells that are distant from blood vessels, as in lens and bone. Moreover, communicating channels are important in development and differentiation. For example, the quiescent uterus transforms to a forcefully contracting organ at the onset of labor; the formation of functional gap junctions at that time creates a syncytium of muscle cells that contract in synchrony.

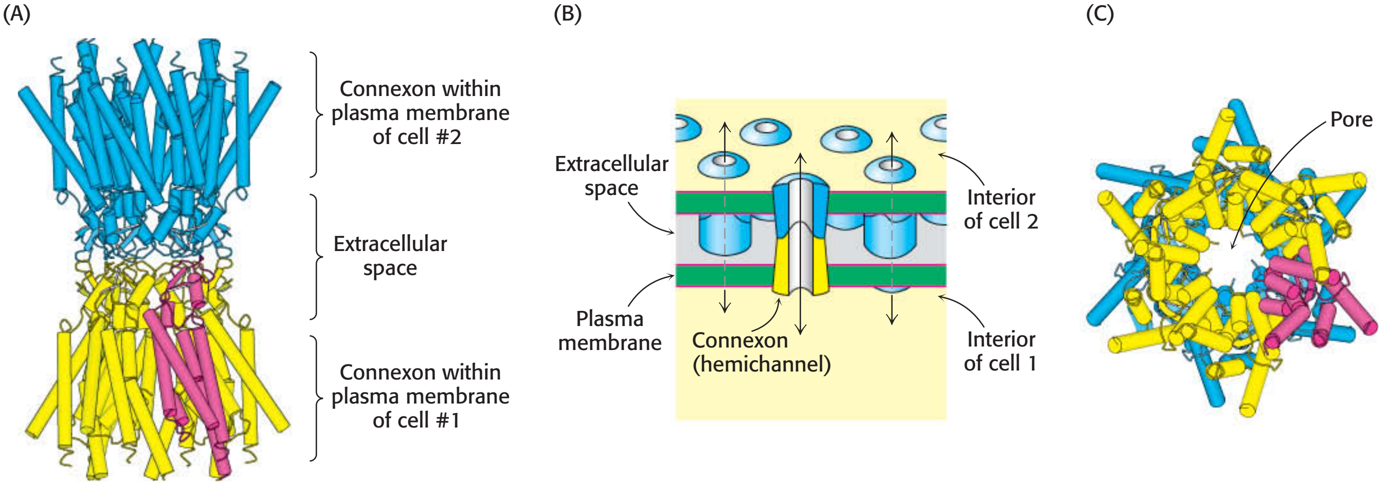


FIGURE 13.32 Structure of a gap junction. (A) Six connexins join to form a connexon, or hemichannel, within the plasma membrane (yellow). A single connexin monomer is highlighted in red. The extracellular region of one connexon binds to the same region of a connexon from another cell (blue), forming a complete gap junction. (B) Schematic view of the gap junction, oriented in the same direction as in (A). (C) A bottom-up view looking through the pore of a gap junction. This perspective is visualized in Figure 13.31. [(A) and (C) Drawn from 2ZW3.pdb; (B) Information from Dr. Werner Loewenstein.]

A cell-to-cell channel is made of 12 molecules of *connexin,* one of a family of transmembrane proteins with molecular masses ranging from 30 to 42 kDa. Each connexin molecule contains four membrane-spanning helices (Figure 13.32A). Six connexin molecules are hexagonally arrayed to form a half-channel, called a *connexon* or *hemichannel*. Two connexons join end to end in the intercellular space to form a functional channel between the communicating cells (Figure 13.32B). Each connexon adopts a funnel-shape: at the cytoplasmic face, the inner diameter of the channel is 35 Å, while at its innermost point, the pore narrows to a diameter of 14 Å (Figure 13.32C). Cell-to-cell channels differ from other membrane channels in three respects: (1) they traverse *two* membranes rather than one; (2) they connect cytoplasm to cytoplasm, rather than to the extracellular space or the lumen of an organelle; and (3) the connexons forming a channel are synthesized by different cells. Gap junctions form readily when cells are brought together. A cell-to-cell channel, once formed, tends to stay open for seconds to minutes. They are closed by high concentrations of calcium ion and by low pH. *The closing of gap junctions by* Ca^{2+} *and* H^{+} *serves to seal normal cells from injured or dying neighbors.* Gap junctions are also controlled by membrane potential and by hormone-induced phosphorylation.

The human genome encodes 21 distinct connexins. Different members of this family are expressed in different tissues. For example, connexin 26 is expressed in key tissues in the ear. Mutations in this connexin are associated with hereditary deafness. The mechanistic basis for this deafness appears to be insufficient transport of ions or second-messenger molecules, such as inositol trisphosphate, between sensory cells.

13.6 Specific Channels Increase the Permeability of Some Membranes to Water

One more important class of channels does not take part in ion transport at all. Instead, these channels increase the rate at which water flows through membranes. As noted in Section 12.3, membranes are reasonably permeable to water. Why, then, are water-specific channels required? In certain tissues, in some circumstances, rapid water transport through membranes is necessary. In the kidney, for example, water must be rapidly reabsorbed into the bloodstream after filtration. Similarly, in the secretion of saliva and tears, water must flow quickly through membranes. These observations suggested the existence of specific water channels, but initially the channels could not be identified.

The channels (now called *aquaporins*) were discovered serendipitously. Peter Agre noticed a protein present at high levels in red-blood-cell membranes that had been missed because the protein does not stain well with Coomassie blue. In addition to red blood cells, this protein was found in large quantities in tissues such as the kidney and the cornea, precisely the tissues thought to contain water channels. On the basis of this observation, further studies were designed, revealing that this 24-kDa membrane protein is, indeed, a water channel.

The structure of aquaporin has been determined (Figure 13.33). The protein consists of six membrane-spanning α helices. Two loops containing hydrophilic residues line the actual channel. Water molecules pass through in single file at a rate of 10^6 molecules per second. Importantly, specific positively charged residues toward the center of the channel prevent the transport of protons through aquaporin. Thus, aquaporin channels will not disrupt proton gradients, which play fundamental roles in energy transduction, as we will see in Chapter 18. Remarkably, the aquaporins are channels that have evolved specifically to conduct uncharged substrates.

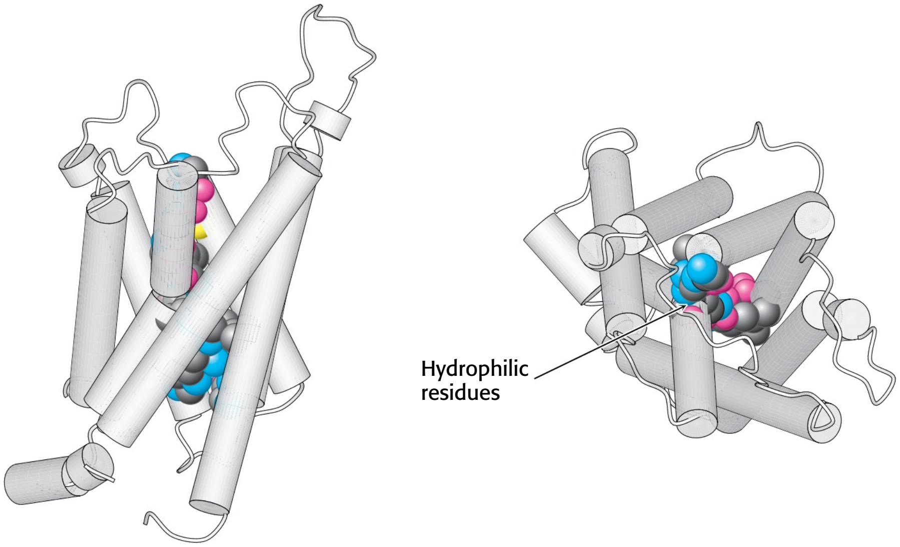


FIGURE 13.33 Structure of aquaporin. The structure of aquaporin viewed from the side (left) and from the top (right). *Notice* the hydrophilic residues (shown as space-filling models) that line the water channel. [Drawn from 1J4N.pdb.]

SUMMARY

13.1 The Transport of Molecules Across a Membrane May Be Active or Passive

For a net movement of molecules across a membrane, two features are required: (1) the molecule must be able to cross a hydrophobic barrier and (2) an energy source must power the movement. Lipophilic molecules can pass through a membrane's hydrophobic interior by simple diffusion. These molecules will move down their concentration gradients. Polar or charged molecules require proteins to form passages through the hydrophobic barrier. Passive transport or facilitated diffusion takes place when an ion or polar molecule moves down its concentration gradient. If a molecule moves against a concentration gradient, an external energy source is required; this movement is referred to as active transport and results in the generation of concentration gradients. The electrochemical potential measures the combined ability of a concentration gradient and an uneven distribution of charge to drive species across a membrane.

13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions and Molecules Across Membranes

Active transport is often carried out at the expense of ATP hydrolysis. P-type ATPases pump ions against a concentration gradient and become transiently phosphorylated on an aspartic acid residue in the process of transport. P-type ATPases, which include the sarcoplasmic reticulum Ca^{2+} ATPase and the Na^+–K^+ ATPase, are integral membrane proteins with conserved structures and catalytic mechanisms. Membrane proteins containing ATP-binding cassette domains are another family of ATP-dependent pumps. Each pump includes four major domains: two domains span the membrane and two others contain ABC P-loop ATPase structures. These pumps are not phosphorylated during pumping; rather, they use the energy of ATP binding and hydrolysis to drive conformational changes that result in the transport of specific substrates across membranes. The multidrug-resistance proteins confer resistance on cancer cells by pumping chemotherapeutic drugs out of a cancer cell before the drugs can exert their effects.

13.3 Lactose Permease is an Archetype of Secondary Transporters That Use One Concentration Gradient to Power the Formation of Another

Carriers are proteins that transport ions or molecules across the membrane without hydrolysis of ATP. They can be classified as uniporters, antiporters, and symporters. Uniporters transport a substrate in either direction, determined by the concentration gradient. Antiporters and symporters can mediate secondary active transport by coupling the uphill flow of one substrate to the downhill flow of another. Antiporters couple the downhill flow of one substrate in one direction to the uphill flow of another in the opposite direction. Symporters move both substrates in the same direction. Studies of the lactose permease from *E. coli* have been a source of insight into both the structures and the mechanisms of secondary transporters.

13.4 Specific Channels Can Rapidly Transport Ions Across Membranes

Ion channels allow the rapid movement of ions across the hydrophobic barrier of the membrane. The activity of individual ion-channel molecules can be observed by using patch-clamp techniques. Many ion channels have a common structural framework. In regard to K^+ channels, hydrated potassium ions must transiently lose their coordinated water molecules as they move to the narrowest part of the channel, termed the selectivity filter. In the selectivity filter, peptide carbonyl groups coordinate the ions. Rapid ion flow through the selectivity filter is facilitated by ion–ion repulsion, with one ion pushing the next ion through the channel. Some ion channels are voltage gated: changes in membrane potential induce conformational changes that open these channels. Many channels spontaneously inactivate after having been open for a short period of time. In some cases, inactivation is due to the binding of a domain of the channel termed the "ball" in the mouth of the channel to block it. Other channels, typified by the acetylcholine receptor, are opened or closed by the binding of ligands. Ligand-gated and voltage-gated channels work in concert to generate action potentials. Inherited mutations or drugs that interfere with the ion channels that produce the action potential can result in potentially life threatening conditions.

13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells

In contrast with many channels, which connect the cell interior with the environment, gap junctions, or cell-to-cell channels, serve to connect the interiors of contiguous groups of cells. A cell-to-cell channel is composed of 12 molecules of connexin, which associate to form two 6-membered connexons.

13.6 Specific Channels Increase the Permeability of Some Membranes to Water

Some tissues contain proteins that increase the permeability of membranes to water. Each water-channel-forming protein, termed an aquaporin, consists of six membrane-spanning α helices and a central channel lined with hydrophilic residues that allow water molecules to pass in single file. Aquaporins do not transport protons.

KEY TERMS

pump (p. 367)
carrier (p. 367)
channel (p. 367)
active transport (p. 367)
facilitated diffusion (passive transport) (p. 367)
ATP-driven pump (p. 367)
primary active transport (p. 367)
secondary active transport (p. 367)
simple diffusion (p. 368)
electrochemical potential (membrane potential) (p. 369)
Na^+–K^+ pump (Na^+–K^+ATPase) (p. 370)
sarcoplasmic reticulum Ca^{2+} ATPase (SERCA) (p. 370)
gastric H^+–K^+ ATPase (p. 370)
P-type ATPase (p. 370)
eversion (p. 372)
cardiotonic steroid (p. 373)
digitalis (p. 373)
multidrug resistance (p. 374)
multidrug-resistance (MDR) protein (P-glycoprotein) (p. 374)
ATP-binding cassette (ABC) domain (p. 374)
ABC transporter (p. 374)
secondary transporter (cotransporter) (p. 376)
antiporter (p. 376)
symporter (p. 376)
uniporter (p. 376)
lactose permease (p. 377)
ion channel (p. 378)
nerve impulse (p. 378)
action potential (p. 378)
patch-clamp technique (p. 379)
gigaseal (p. 379)
selectivity filter (p. 381)
voltage-gated channel (p. 383)
ball-and-chain model (p. 384)
neurotransmitter (p. 385)
acetylcholine (p. 385)
synaptic cleft (p. 385)
acetylcholine receptor (p. 385)
ligand-gated channel (p. 386)
equilibrium potential (p. 387)
Nernst equation (p. 387)
long QT syndrome (LQTS) (p. 388)
gap junction (cell-to-cell channels) (p. 389)
connexin (p. 390)
connexon (hemichannel) (p. 390)
aquaporin (p. 391)

PROBLEMS

1. *A helping hand.* Differentiate between simple diffusion and facilitated diffusion.

2. *Powering movement.* What are the two forms of energy that can power active transport?

3. *Carriers.* Name the three types of carrier proteins. Which of these can mediate secondary active transport?

4. *The price of extrusion.* What is the free-energy cost at 25°C of pumping Ca^{2+} out of a cell when the cytoplasmic concentration is 0.4 μM, the extracellular concentration is 1.5 mM, and the membrane potential is −60 mV?

5. *Equilibrium potentials.* For a typical mammalian cell, the intracellular and extracellular concentrations of the chloride ion (Cl^-) are 4 mM and 150 mM, respectively.

For the calcium ion (Ca^{2+}), the intracellular and extracellular concentrations are 0.2 μM and 1.8 mM, respectively. Calculate the equilibrium potentials at 37°C for these two ions.

6. *How sweet it is*. Some animal cells take up glucose by a symporter powered by the simultaneous entry of Na^+. The entry of Na^+ provides a free-energy input of 10.8 $kJ\ mol^{-1}$ (2.6 $kcal\ mol^{-1}$) under typical cellular conditions (external $[Na^+] = 143$ mM, internal $[Na^+] =$ 14 mM, and membrane potential $= -50$ mV). How large a concentration gradient of glucose at 37°C can be generated by this free-energy input?

7. *Variations on a theme*. Write a detailed mechanism for transport by the Na^+–K^+ ATPase based on analogy with the mechanism of the Ca^{2+} ATPase shown in Figure 13.5.

8. *Pumping protons*. Design an experiment to show that the action of lactose permease can be reversed in vitro to pump protons.

9. *Opening channels*. Differentiate between ligand-gated and voltage-gated channels.

10. *Different directions*. The K^+ channel and the Na^+ channel have similar structures and are arranged in the same orientation in the cell membrane. Yet the Na^+ channel allows sodium ions to flow into the cell and the K^+ channel allows potassium ions to flow out of the cell. Explain.

11. *Differing mechanisms*. Distinguish the mechanisms by which uniporters and channels transport ions or molecules across the membrane.

12. *Short circuit*. Carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) is a proton ionophore: it enables protons to pass freely through membranes. Treatment of *E. coli* with FCCP prevents the accumulation of lactose in these cells. Explain.

13. *Working together*. The human genome contains more than 20 connexin-encoding genes. Several of these genes are expressed in high levels in the heart. Why are connexins so highly expressed in cardiac tissue?

14. *Structure–activity relations*. On the basis of the structure of tetrodotoxin, propose a mechanism by which the toxin inhibits Na^+ flow through the Na^+ channel.

15. *Hot stuff*. When SERCA is incubated with [γ-^{32}P]ATP (a form of ATP in which the terminal phosphate is labeled with radioactive ^{32}P) and calcium at 0°C for 20 seconds and analyzed by gel electrophoresis, a radioactive band is observed at the molecular weight corresponding to full-length SERCA. Why is a labeled band observed? Would you expect a similar band if you were performing a similar assay, with a suitable substrate, for the MDR protein?

16. *A dangerous snail*. Cone snails are carnivores that inject a powerful set of toxins into their prey, leading to rapid paralysis. Many of these toxins are found to bind to specific ion-channel proteins. Why are such molecules so toxic? How might such toxins be useful for biochemical studies?

17. *Pause for effect*. Immediately after the repolarization phase of an action potential, the neuronal membrane is temporarily unable to respond to the stimulation of a second action potential, a phenomenon referred to as the *refractory period*. What is the mechanistic basis for the refractory period?

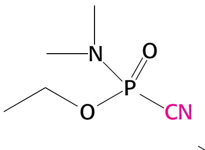
18. *Only a few*. Why do only a small number of sodium ions need to flow through the Na^+ channel to change the membrane potential significantly?

19. *More than one mechanism*. How might a mutation in a cardiac voltage-dependent sodium channel cause long QT syndrome?

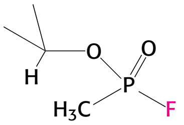
20. *Mechanosensitive channels*. Many species contain ion channels that respond to mechanical stimuli. On the basis of the properties of other ion channels, would you expect the flow of ions through a single open mechanosensitive channel to increase in response to an appropriate stimulus? Why or why not?

21. *Concerted opening*. Suppose that a channel obeys the concerted allosteric model (MWC model, Section 7.2). The binding of ligand to the R state (the open form) is 20 times as tight as that to the T state (the closed form). In the absence of ligand, the ratio of closed to open channels is 10^5. If the channel is a tetramer, what is the fraction of open channels when 1, 2, 3, and 4 ligands are bound?

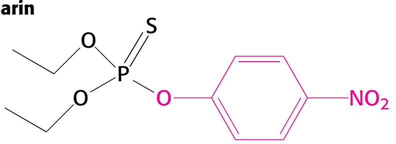
22. *Respiratory paralysis*. The neurotransmitter acetylcholine is degraded by a specific enzyme that is inactivated by Tabun, sarin, and parathion. On the basis of the structures below, propose a possible basis for their lethal actions.



Tabun



Sarin



Parathion

23. *Ligand-induced channel opening*. The ratio of open to closed forms of the acetylcholine receptor channel containing zero, one, and two bound acetylcholine molecules is 5×10^{-6}, 1.2×10^{-3}, and 14, respectively.

(a) By what factor is the open-to-closed ratio increased by the binding of the first acetylcholine molecule? The second acetylcholine molecule?

(b) What are the corresponding free-energy contributions to channel opening at 25°C?

(c) Can the allosteric transition be accounted for by the MWC concerted model (Section 7.2)?

24. *Frog poison*. Batrachotoxin (BTX) is a steroidal alkaloid from the skin of *Phyllobates terribilis*, a poisonous Colombian frog (the source of the poison used on blowgun darts). In the presence of BTX, Na^+ channels in an excised patch stay persistently open when the membrane is depolarized. They close when the membrane is repolarized. Which transition is blocked by BTX?

25. *Valium target*. γ-Aminobutyric acid (GABA) opens channels that are specific for chloride ions. The $GABA_A$ receptor channel is pharmacologically important because it is the target of Valium, which is used to diminish anxiety.

(a) The extracellular concentration of Cl^- is 123 mM and the intracellular concentration is 4 mM. In which direction does Cl^- flow through an open channel when the membrane potential is in the -60 mV to $+30$ mV range?

(b) What is the effect of Cl^--channel opening on the excitability of a neuron?

(c) The hydropathy profile of the $GABA_A$ receptor resembles that of the acetylcholine receptor. Predict the number of subunits in this Cl^- channel.

26. *Understanding SERCA*. To study the mechanism of SERCA, you prepare membrane vesicles containing this protein oriented such that its ATP binding site is on the outer surface of the vesicle. To measure pump activity, you use an assay that detects the formation of inorganic phosphate in the medium. When you add calcium and ATP to the medium, you observe phosphate production for only a short period of time. Only after the addition of calcimycin, a molecule that makes membranes selectively permeable to calcium, do you observe sustained phosphate production. Explain.

Chapter Integration Problem

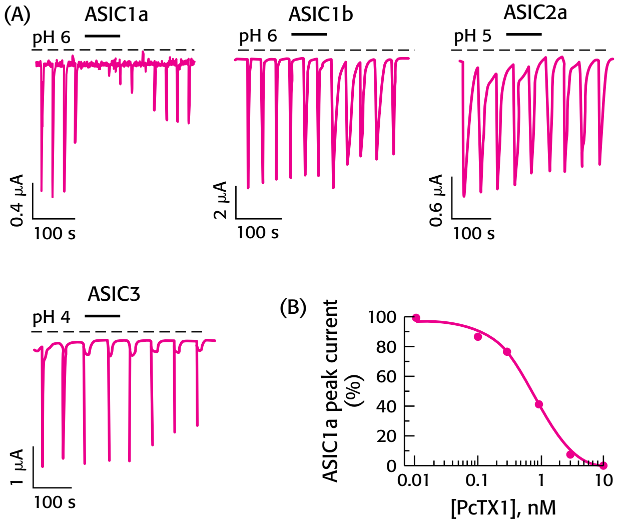
27. *Speed and efficiency matter*. Acetylcholine is rapidly destroyed by the enzyme acetylcholinesterase. This enzyme, which has a turnover number of 25,000 per second, has attained catalytic perfection with a k_{cat}/K_M of $2 \times 10^8\ M^{-1}s^{-1}$. Why is the efficiency of this enzyme physiologically crucial?

Mechanism Problem

28. *Remembrance of mechanisms past*. Acetylcholinesterase converts acetylcholine into acetate and choline. Like serine proteases, acetylcholinesterase is inhibited by DIPF. Propose a catalytic mechanism for acetylcholine digestion by acetylcholinesterase. Show the reaction as chemical structures.

Data Interpretation Problems

29. *Tarantula toxin*. Acid sensing is associated with pain, tasting, and other biological activities (Chapter 33). Acid sensing is carried out by a ligand-gated channel that permits Na^+ influx in response to H^+. This family of acid-sensitive ion channels (ASICs) includes a number of members. Psalmotoxin 1 (PcTX1), a venom from the tarantula, inhibits some members of this family. The following electrophysiological recordings of cells containing several members of the ASIC family were made in the presence of the toxin at a concentration of 10 nM. The channels were opened by changing the pH from 7.4 to the indicated values. The PcTX1 was present for a short time (indicated by the black bar above the recordings below), after which time it was rapidly washed from the system.



(A) Electrophysiological recordings of cells exposed to tarantula toxin. (B) Plot of peak current of a cell containing the ASIC1a protein versus the toxin concentration. [Data from P. Escoubas et al., *J. Biol. Chem.* 275:25116–25121, 2000.]

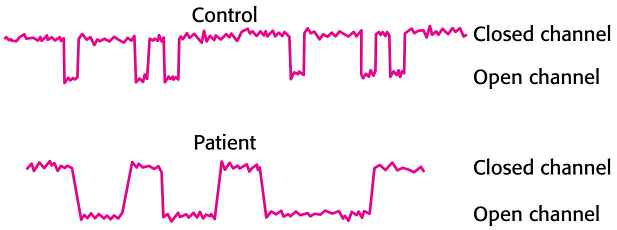
(a) Which member of the ASIC family—ASIC1a, ASIC1b, ASIC2a, or ASIC3—is most sensitive to the toxin?

(b) Is the effect of the toxin reversible? Explain.

(c) What concentration of PcTX1 yields 50% inhibition of the sensitive channel?

30. *Channel problems 1*. A number of pathological conditions result from mutations in the acetylcholine receptor

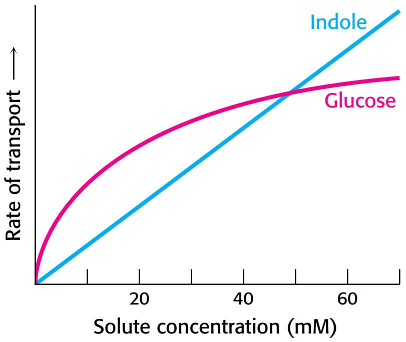
channel. One such mutation in the β subunit, βV266M, causes muscle weakness and rapid fatigue. An investigation of the acetylcholine-generated currents through the acetylcholine receptor channel for both a control and a patient yielded the following results.



What is the effect of the mutation on channel function? Suggest some possible biochemical explanations for the effect.

31. *Channel problems 2.* The acetylcholine receptor channel can also undergo mutation leading to fast-channel syndrome (FCS), with clinical manifestations similar to those of slow-channel syndrome (Problem 30). What would the recordings of ion movement look like in this syndrome? Suggest a biochemical explanation.

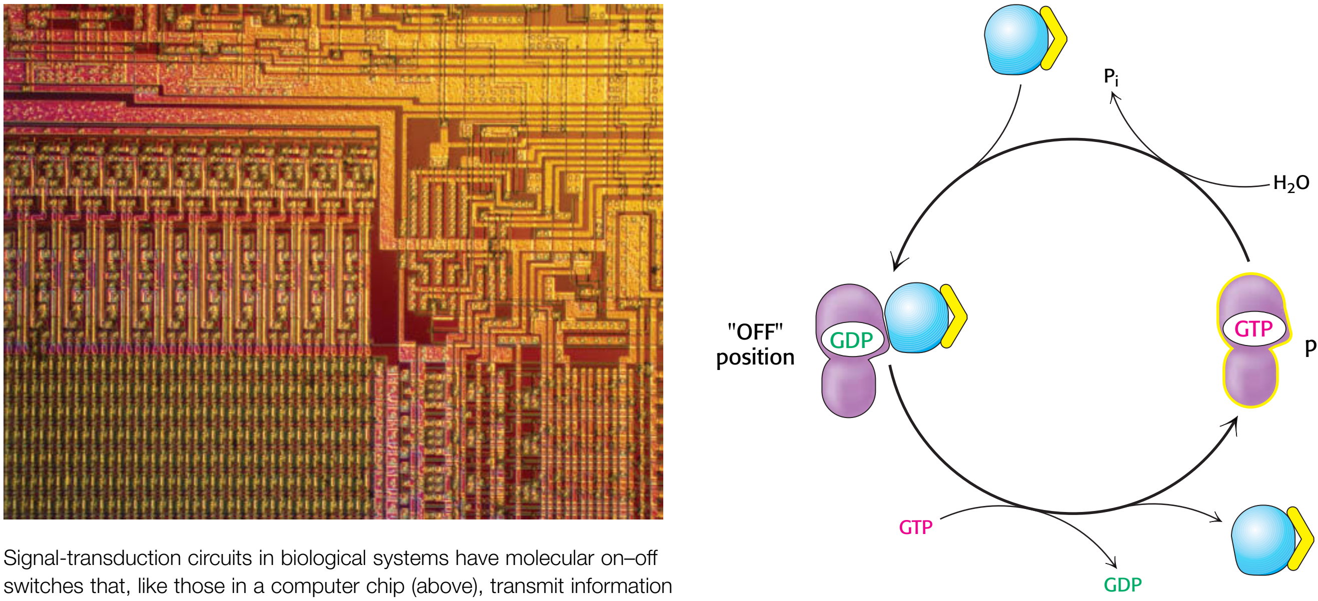
32. *Transport differences.* The rate of transport of two molecules, indole and glucose, across a cell membrane is shown below. What are the differences between the transport mechanisms of the two molecules? Suppose that ouabain inhibited the transport of glucose. What would this inhibition suggest about the mechanism of transport?



CHAPTER

14

Signal-Transduction Pathways



Signal-transduction circuits in biological systems have molecular on–off switches that, like those in a computer chip (above), transmit information when "on." Common among these circuits are those including G proteins (right), which transmit a signal when bound to GTP and are silent when bound to GDP. [(Left) Astrid & Hanns-Frieder Michler/Science Source.]

A cell is highly responsive to specific chemicals in its environment: it may adjust its metabolism or alter gene-expression patterns on sensing the presence of these molecules. In multicellular organisms, these chemical signals are crucial to coordinating physiological responses (Figure 14.1). Three examples of molecular signals that stimulate a physiological response are epinephrine (sometimes called adrenaline), insulin, and epidermal growth factor (EGF). When a mammal is threatened, its adrenal glands release the hormone epinephrine, which stimulates the mobilization of energy stores and leads to improved cardiac function. After a full meal, the β cells in the pancreas release insulin, which stimulates a host of physiological responses, including the uptake of glucose from the bloodstream and its storage as glycogen. The release of EGF in response to a wound stimulates specific cells to grow and divide. In all these cases, the cell receives information that a certain molecule within its environment is present above some threshold concentration. The chain of events that converts the message "this molecule is present" into the ultimate physiological response is called *signal transduction.*

Signal-transduction pathways often comprise many components and branches. They can thus be immensely complicated and confusing. However, the logic of signal transduction can be simplified by examining the common strategies and classes of molecules that recur in these pathways.

OUTLINE

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes

14.3 EGF Signaling: Signal-Transduction Systems Are Poised to Respond

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases

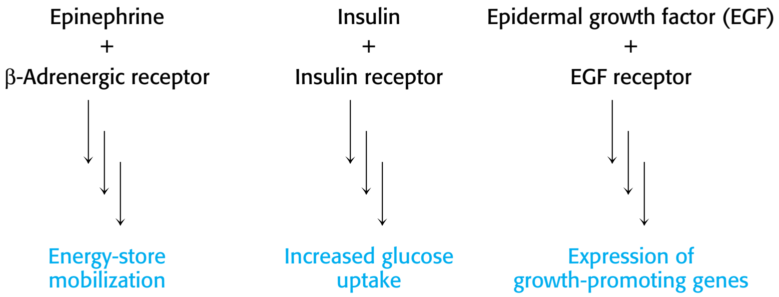


FIGURE 14.1 Three signal-transduction pathways. The binding of signaling molecules to their receptors initiates pathways that lead to important physiological responses.

These principles are introduced here because signal-transduction pathways affect essentially all of the metabolic pathways that we will be exploring throughout the rest of the book.

Signal transduction depends on molecular circuits

Signal-transduction pathways follow a broadly similar course that can be viewed as a molecular circuit (Figure 14.2). All such circuits contain certain key steps:

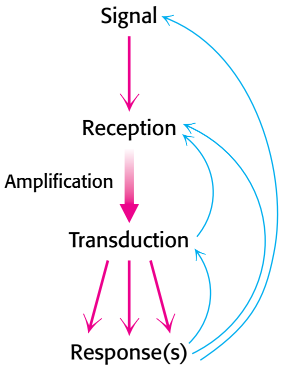


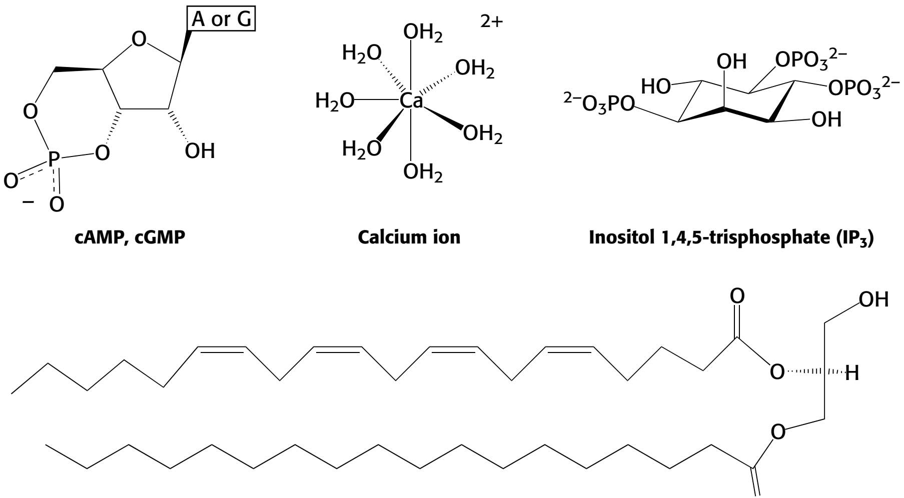
FIGURE 14.2 Principles of signal transduction. An environmental signal is first received by interaction with a cellular component, most often a cell-surface receptor. The information that the signal has arrived is then converted into other chemical forms, or *transduced*. Typically, the transduction process comprises many steps. The signal is often amplified before evoking a response. Feedback pathways regulate the entire signaling process.

1. *Release of the Primary Messenger*. A stimulus such as a wound or digested meal triggers the release of the signal molecule, also called the *primary messenger*.

2. *Reception of the Primary Messenger*. Most signal molecules do not enter cells. Instead, proteins in the cell membrane act as *receptors* that bind the signal molecules and transfer the information that the molecule has bound from the external environment to the cell's interior. Receptors span the cell membrane and thus have both extracellular and intracellular components. A binding site on the extracellular side specifically recognizes the signal molecule (often referred to as the *ligand*). Such binding sites are analogous to enzyme active sites except that no catalysis takes place within them. The interaction of the ligand and the receptor alters the tertiary or quaternary structure of the receptor so as to induce a structural change on the intracellular side.

3. *Delivery of the Message Inside the Cell by the Second Messenger*. Other small molecules, called *second messengers,* are used to relay information from receptor–ligand complexes. Second messengers are intracellular molecules that change in concentration in response to environmental signals and mediate the next step in the molecular information circuit. Some particularly important second messengers are cyclic AMP (cAMP) and cyclic GMP (cGMP), calcium ion, inositol 1,4,5-trisphosphate (IP_3), and diacylglycerol (DAG; Figure 14.3).

The use of second messengers has several consequences. First, the signal may be amplified significantly: only a small number of receptor molecules may be activated by the direct binding of signal molecules, but each activated receptor molecule can lead to the generation of many second messengers. Thus, *a low concentration of signal in the environment, even as little as a single molecule, can yield a large intracellular signal and response.* Second, these messengers are often free to diffuse to other cellular compartments where they can influence processes throughout the cell. Third, the use of common second messengers in multiple signaling pathways creates both opportunities and potential problems. Input from several signaling pathways, often called *cross talk,* may alter the concentration of a common second messenger. Cross talk permits more finely tuned regulation of cell activity than



cAMP, cGMP **Calcium ion** **Inositol 1,4,5-trisphosphate (IP_3)**

Diacylglycerol (DAG)

FIGURE 14.3 Common second messengers. Second messengers are intracellular molecules that change in concentration in response to environmental signals. That change in concentration conveys information inside the cell.

would the action of individual independent pathways. However, inappropriate cross talk can result in the misinterpretation of changes in second-messenger concentration.

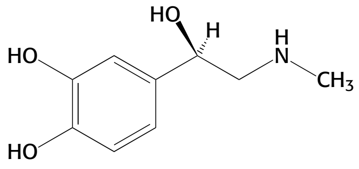
4. *Activation of Effectors That Directly Alter the Physiological Response.* The ultimate effect of the signal pathway is to activate (or inhibit) the pumps, channels, enzymes, and transcription factors that directly control metabolic pathways, gene expression, and the permeability of membranes to specific ions.

5. *Termination of the Signal.* After a cell has completed its response to a signal, the signaling process must be terminated or the cell loses its responsiveness to new signals. Moreover, signaling processes that fail to terminate properly can have highly undesirable consequences. As we will see, many cancers are associated with signal-transduction processes that are not properly terminated, especially processes that control cell growth.

In this chapter, we will examine components of the three signal-transduction pathways shown in Figure 14.1. In doing so, we will see several classes of adaptor domains present in signal-transduction proteins. These domains usually recognize specific classes of molecules and help transfer information from one protein to another. The components described in the context of these three pathways recur in many other signal-transduction pathways; bear in mind that the specific examples are representative of many such pathways.

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves

Epinephrine is a hormone secreted by the adrenal glands of mammals in response to internal and external stressors. It exerts a wide range of effects—referred to as the *fight-or-flight response*—to help organisms anticipate the need for rapid muscular activity, including acceleration of heart rate, dilation of the smooth muscle of the airways, and initiation of the breakdown of glycogen (Section 21.3) and fatty acids (Section 22.2). Epinephrine signaling begins with ligand binding to a protein called the *β-adrenergic receptor* (β-AR). The β-AR is a member of the largest class of cell-surface receptors, called the *seven-transmembrane-helix* (7TM) *receptors*. Members of this family are responsible for transmitting information initiated by signals as



Epinephrine

TABLE 14.1 Biological functions mediated by 7TM receptors

Hormone action
Hormone secretion
Neurotransmission
Chemotaxis
Exocytosis
Control of blood pressure
Embryogenesis
Cell growth and differentiation
Development
Smell
Taste
Vision
Viral infection

Information from J. S. Gutkind, *J. Biol. Chem.* 273:1839–1842, 1998.

diverse as hormones, neurotransmitters, odorants, tastants, and even photons (Table 14.1). More than 20,000 such receptors are now known, including nearly 800 encoded in the human genome. Furthermore, about one-third of the marketed therapeutic drugs target receptors of this class. As the name indicates, these receptors contain seven helices that span the membrane bilayer (Figure 14.4).

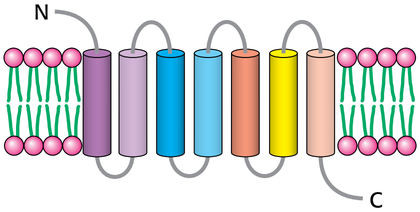


FIGURE 14.4 The 7TM receptor. Schematic representation of a 7TM receptor showing its passage through the membrane seven times.

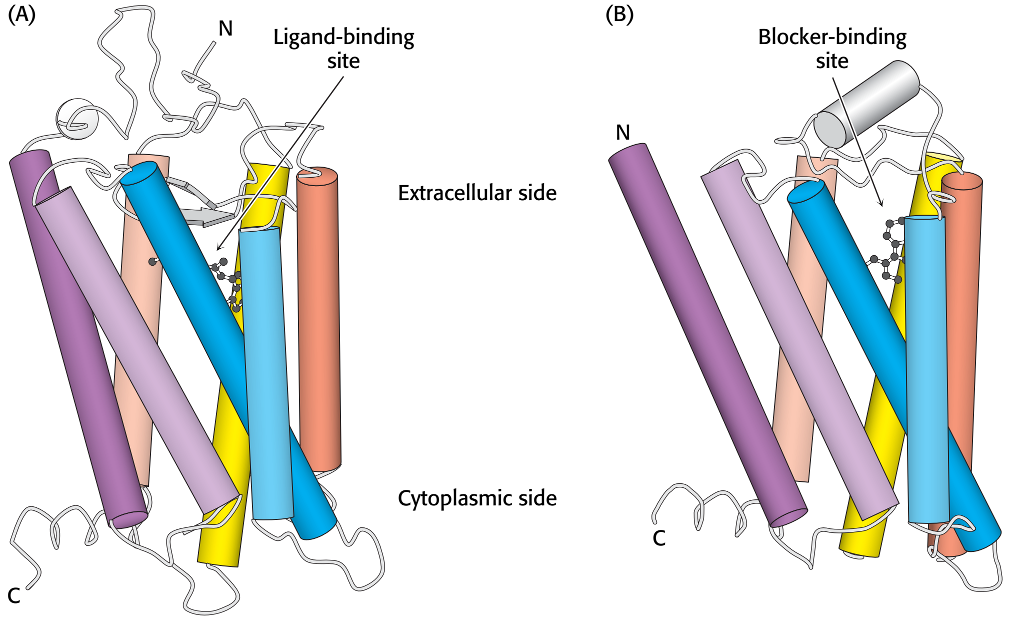
The first member of the 7TM receptor family to have its three-dimensional structure determined was *rhodopsin* (Figure 14.5A), a protein in the retina of the eye that senses the presence of photons and initiates the signaling cascade responsible for visual sensation. A single lysine residue within rhodopsin is covalently modified by a form of vitamin A, 11-*cis*-retinal. This modification is located near the extracellular side of the receptor, within the region surrounded by the seven transmembrane helices. As will be considered in greater detail in Section 33.3, exposure to light induces the isomerization of 11-*cis*-retinal to its all-*trans* form, producing a structural change in the receptor that results in the initiation of an action potential that is ultimately interpreted by the brain as visual stimulus.

In 2007, the first three-dimensional structure of the β_2 subtype of the human adrenergic receptor (β_2-AR) bound to an inhibitor was solved by x-ray crystallography. This inhibitor, carazolol, competes with epinephrine for binding to the β_2-AR, much in the same way that competitive inhibitors act at enzyme active sites (Section 8.5). The structure of the β_2-AR revealed considerable similarities with that of rhodopsin, particularly with respect to the locations of 11-*cis*-retinal in rhodopsin and the binding site for carazolol (Figure 14.5B).

Ligand binding to 7TM receptors leads to the activation of heterotrimeric G proteins

What is the next step in the pathway? The conformational change in the receptor's cytoplasmic domain activates a protein called a *G protein*, named for the fact that it binds guanyl nucleotides. The activated G protein stimulates the activity of adenylate cyclase, an enzyme that catalyzes the conversion of ATP into cAMP. The G protein and adenylate cyclase remain attached to the membrane, whereas cAMP, a second messenger, can travel

FIGURE 14.5 Structures of rhodopsin and the β_2-adrenergic receptor. Three-dimensional structure of rhodopsin (A) and the β_2-adrenergic receptor (β_2-AR). (B). *Notice* the resemblance in the overall architecture of both receptors and the similar locations of the rhodopsin ligand 11-*cis*-retinal and the β_2-AR blocker carazolol. [Drawn from 1F88.pdb and 2RH1.pdb.]



Rhodopsin

β_2-adrenergic receptor

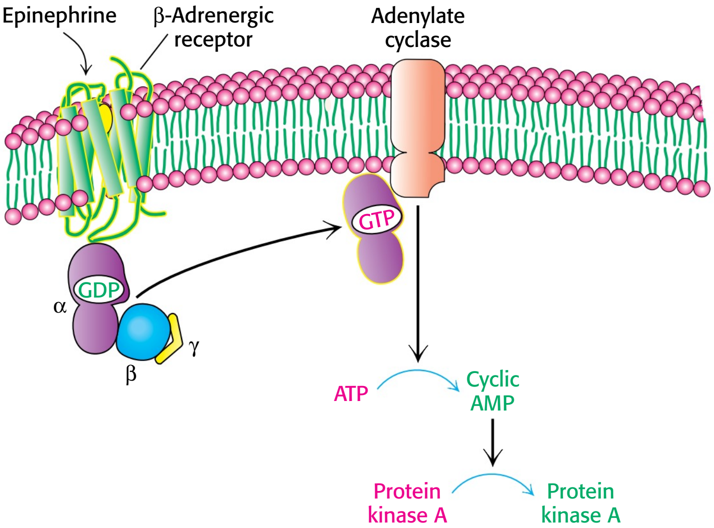


FIGURE 14.6 Activation of protein kinase A by a G-protein pathway. Hormone binding to a 7TM receptor initiates a signal-transduction pathway that acts through a G protein and cAMP to activate protein kinase A.

throughout the cell carrying the signal originally brought by the binding of epinephrine. Figure 14.6 provides a broad overview of these steps.

Let us consider the role of the G protein in this signaling pathway in greater detail. In its unactivated state, the G protein is bound to GDP. In this form, the G protein exists as a heterotrimer consisting of α, β, and γ subunits; the α subunit (referred to as G_α) binds the nucleotide (Figure 14.7). The α subunit is a member of the P-loop NTPase family (Section 9.4), and the P-loop participates in nucleotide binding. The α and γ subunits are usually anchored to the membrane by covalently attached fatty acids. *The role of the hormone-bound receptor is to catalyze the exchange of GTP for bound GDP.* The interaction between the hormone–β_2-AR complex and the heterotrimeric G protein was illustrated in molecular detail when the crystal structure of this complex was determined in 2011. In this structure, a synthetic *agonist*, or small molecule that activates a receptor, was used to induce the active conformation of the β_2-AR. Agonist binding results in the movement of two transmembrane helices, yielding an extensive interaction surface for the G_α subunit of the

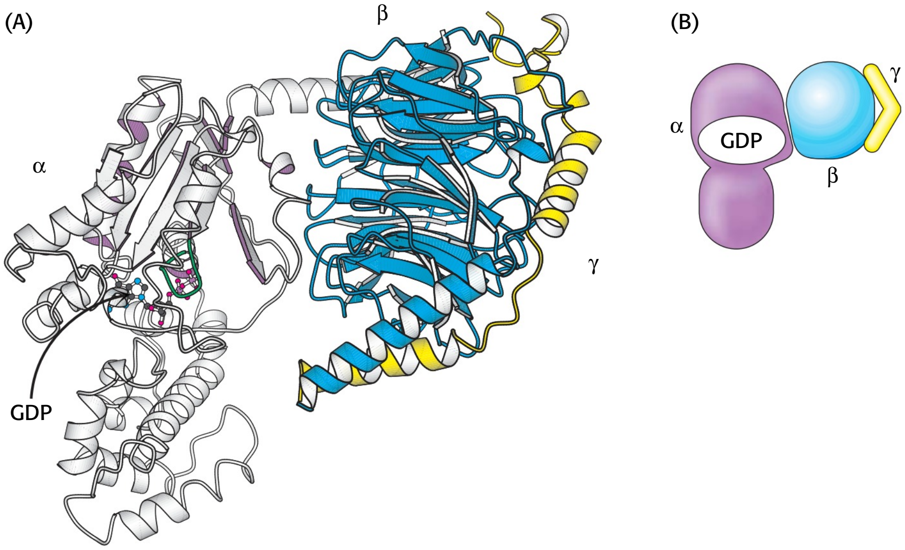


FIGURE 14.7 A heterotrimeric G protein. (A) A ribbon diagram shows the relation between the three subunits. In this complex, the α subunit (gray and purple) is bound to GDP. *Notice* that GDP is bound in a pocket close to the surface at which the α subunit interacts with the βγ dimer. (B) A schematic representation of the heterotrimeric G protein. [Drawn from 1GOT.pdb.]

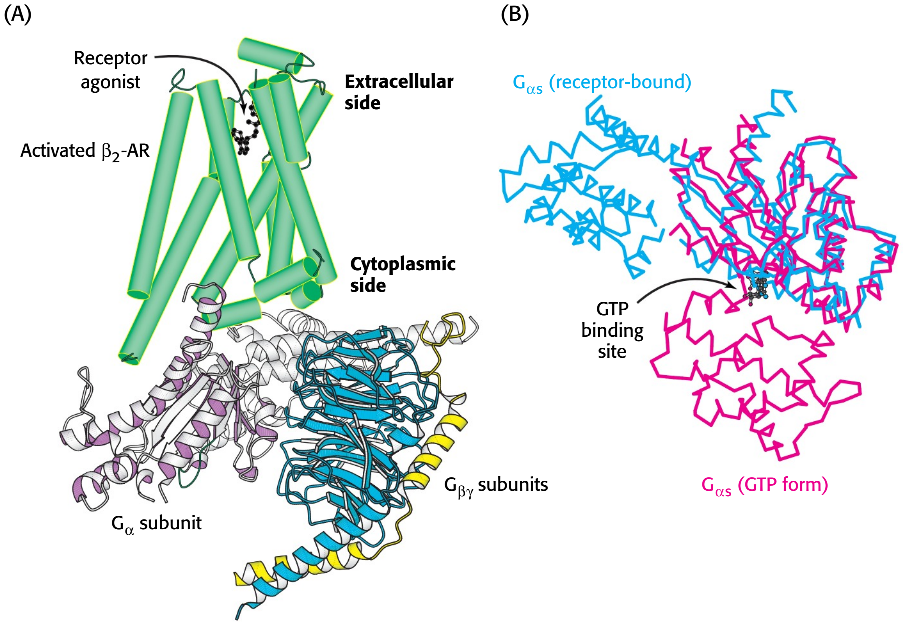


FIGURE 14.8 The complex between the activated β_2-AR and a heterotrimeric G protein. (A) When the β_2-AR (green) binds a receptor agonist, the cytoplasmic face of the receptor forms an interaction surface with the G_α subunit of a heterotrimeric G protein. (B) The interaction with the activated receptor leads to a substantial conformational change in the G_α protein, in which the GTP binding site is opened, enabling nucleotide exchange. In this figure, G_α in its GTP form is shown in red and in its receptor-bound form is shown in blue. [Drawn from 3SN6.pdb and 1AZT.pdb.]

heterotrimer (Figure 14.8A). When bound to the receptor, the nucleotide-binding site of G_α opens substantially, enabling the displacement of GDP by GTP (Figure 14.8B).

On GTP binding, the α subunit simultaneously dissociates from the βγ dimer ($G_{\beta\gamma}$), transmitting the signal that the receptor has bound its ligand. *A single hormone–receptor complex can stimulate nucleotide exchange in many G-protein heterotrimers*. Thus, hundreds of G_α molecules are converted from their GDP form into their GTP form for each bound molecule of hormone, giving an amplified response. Because they signal through G proteins, 7TM receptors are often called *G-protein-coupled receptors* (GPCRs).

Activated G proteins transmit signals by binding to other proteins

In the GTP form, the surface of G_α that had been bound to $G_{\beta\gamma}$ has changed its conformation from the GDP form so that it no longer has a high affinity for $G_{\beta\gamma}$. This surface is now exposed for binding to other proteins. In the β-AR pathway, the new binding partner is *adenylate cyclase,* the enzyme that converts ATP into cAMP. This enzyme is a membrane protein that contains 12 membrane-spanning helices; two large cytoplasmic domains form the catalytic part of the enzyme (Figure 14.9). The interaction of G_α with adenylate cyclase favors a more catalytically active conformation of the enzyme, thus stimulating cAMP production. Indeed, the G_α subunit that participates in the β-AR pathway is called $G_{\alpha s}$ ("s" stands for stimulatory). *The net result is that*

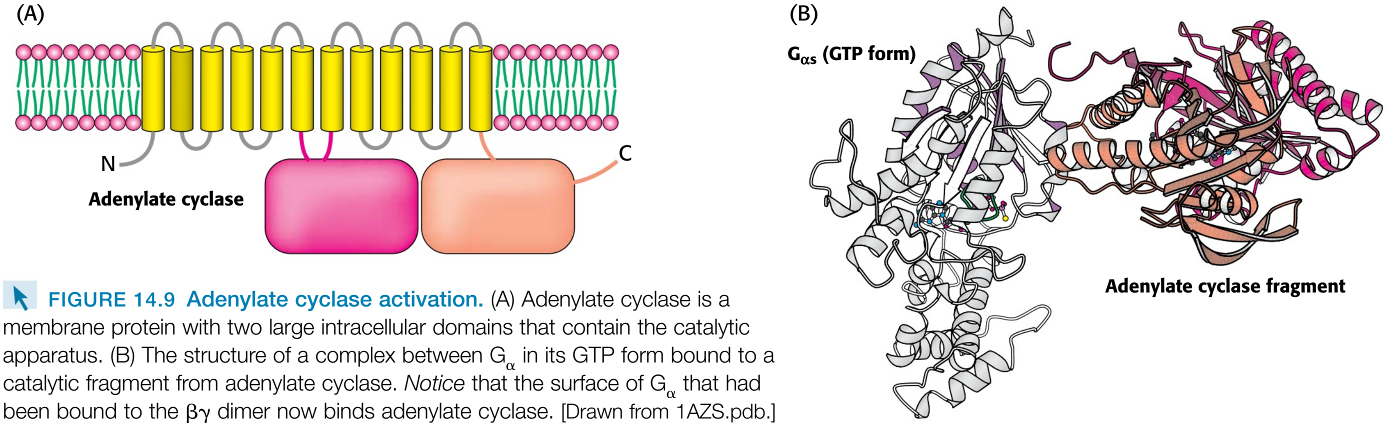


FIGURE 14.9 Adenylate cyclase activation. (A) Adenylate cyclase is a membrane protein with two large intracellular domains that contain the catalytic apparatus. (B) The structure of a complex between G_α in its GTP form bound to a catalytic fragment from adenylate cyclase. *Notice* that the surface of G_α that had been bound to the βγ dimer now binds adenylate cyclase. [Drawn from 1AZS.pdb.]

the binding of epinephrine to the receptor on the cell surface increases the rate of cAMP production inside the cell. The generation of cAMP by adenylate cyclase provides a second level of amplification because each activated adenylate cyclase can convert many molecules of ATP into cAMP.

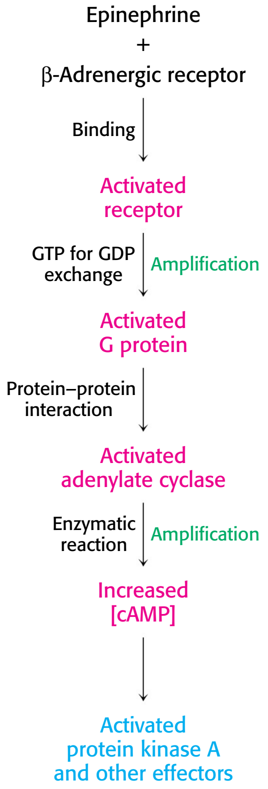


FIGURE 14.10 Epinephrine signaling pathway. The binding of epinephrine to the β-adrenergic receptor initiates the signal-transduction pathway. The process in each step is indicated (in black) at the left of each arrow. Steps that have the potential for signal amplification are indicated at the right in green.

Cyclic AMP stimulates the phosphorylation of many target proteins by activating protein kinase A

The increased concentration of cAMP can affect a wide range of cellular processes. In the muscle, cAMP stimulates the production of ATP for muscle contraction. In other cell types, cAMP enhances the degradation of storage fuels, increases the secretion of acid by the gastric mucosa, leads to the dispersion of melanin pigment granules, diminishes the aggregation of blood platelets, and induces the opening of chloride channels. How does cAMP influence so many cellular processes? *Most effects of cAMP in eukaryotic cells are mediated by the activation of a single protein kinase.* This key enzyme is *protein kinase A* (PKA).

As described earlier, PKA consists of two regulatory (R) chains and two catalytic (C) chains (R_2C_2; see Figure 10.16). In the absence of cAMP, the R_2C_2 complex is catalytically inactive. The binding of cAMP to the regulatory chains releases the catalytic chains, which are catalytically active on their own. Activated PKA then phosphorylates specific serine and threonine residues in many targets to alter their activity. For instance, PKA phosphorylates two enzymes that lead to the breakdown of glycogen, the polymeric store of glucose, and the inhibition of further glycogen synthesis (Section 21.3). *PKA stimulates the expression of specific genes* by phosphorylating a transcriptional activator called the cAMP response element binding (CREB) protein. This activity of PKA illustrates that signal-transduction pathways can extend into the nucleus to alter gene expression.

The signal-transduction pathway initiated by epinephrine is summarized in Figure 14.10.

G proteins spontaneously reset themselves through GTP hydrolysis

How is the signal initiated by epinephrine switched off? *G_α subunits have intrinsic GTPase activity,* which is used to hydrolyze bound GTP to GDP and P_i. This hydrolysis reaction is slow, however, requiring from seconds to minutes. Thus, the GTP form of G_α is able to activate downstream components of the signal-transduction pathway before it is deactivated by GTP hydrolysis. In essence, *the bound GTP acts as a built-in clock that spontaneously resets the G_α subunit after a short time period.* After GTP hydrolysis and the release of P_i, the GDP-bound form of G_α then reassociates with $G_{\beta\gamma}$ to re-form the inactive heterotrimeric protein (Figure 14.11).

The hormone-bound activated receptor must be reset as well to prevent the continuous activation of G proteins. This resetting is accomplished by two processes (Figure 14.12). First, the hormone dissociates, returning the receptor to its initial, unactivated state. The likelihood that the receptor remains in

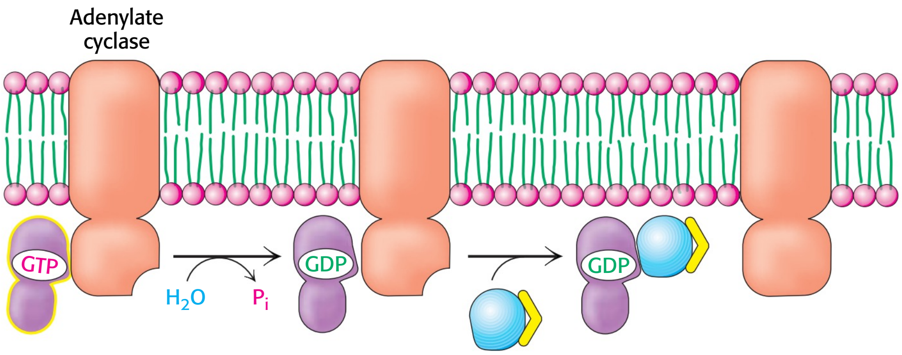


FIGURE 14.11 Resetting G_α. On hydrolysis of the bound GTP by the intrinsic GTPase activity of G_α, G_α reassociates with the βγ dimer to form the heterotrimeric G protein, thereby terminating the activation of adenylate cyclase.

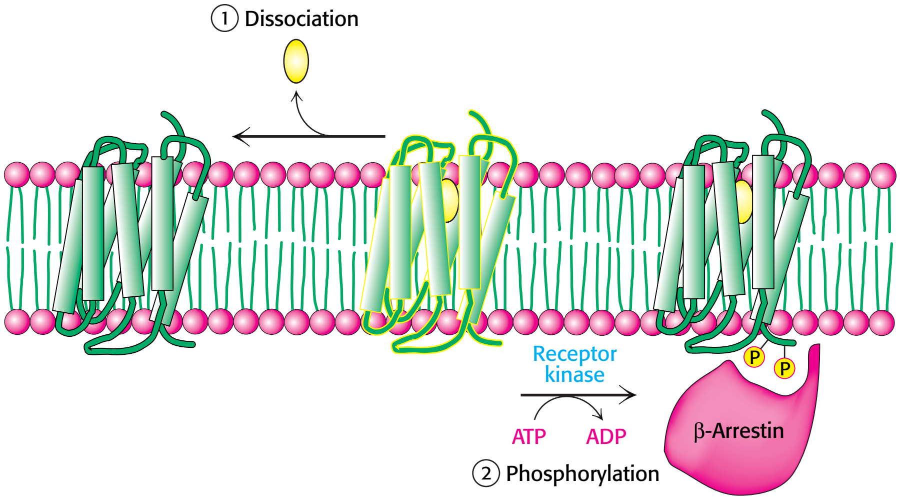


FIGURE 14.12 Signal termination. Signal transduction by the 7TM receptor is halted (1) by dissociation of the signal molecule from the receptor and (2) by phosphorylation of the cytoplasmic C-terminal tail of the receptor and the subsequent binding of β-arrestin.

its unbound state depends on the extracellular concentration of hormone. Second, the signaling cascade initiated by the hormone–receptor complex activates a kinase that phosphorylates serine and threonine residues in the carboxyl-terminal tail of the receptor. These phosphorylation events result in the deactivation of the receptor. In the example under consideration, *β-adrenergic-receptor kinase* (also called *G-protein receptor kinase 2,* or *GRK2*) phosphorylates the carboxyl-terminal tail of the hormone–receptor complex but not the unoccupied receptor. Finally, the molecule *β-arrestin* binds to the phosphorylated receptor and further diminishes its ability to activate G proteins.

Some 7TM receptors activate the phosphoinositide cascade

We now turn to another common second-messenger cascade, also employing a 7TM receptor, that is used by many hormones to evoke a variety of responses. The *phosphoinositide cascade,* like the cAMP cascade, converts extracellular signals into intracellular ones. The intracellular second messengers formed by activation of this pathway arise from the cleavage of *phosphatidylinositol 4,5-bisphosphate* (PIP_2), a phospholipid present in cell membranes. An example of a signaling pathway based on the phosphoinositide cascade is the one triggered by the receptor for angiotensin II, a peptide hormone that controls blood pressure.

Each type of 7TM receptor signals through a distinct G protein. Whereas the β-adrenergic receptor activates the G protein $G_{\alpha s}$, the angiotensin II receptor activates a G protein called $G_{\alpha q}$. In its GTP form, $G_{\alpha q}$ binds to and activates the β isoform of the enzyme *phospholipase C*. This enzyme catalyzes the cleavage of PIP_2 into the two second messengers—inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG; Figure 14.13).

IP_3 is soluble and diffuses away from the membrane. This second messenger causes the rapid release of Ca^{2+} from the intracellular stores in the endoplasmic reticulum (ER), which accumulates a reservoir of Ca^{2+} through the action of transporters such as Ca^{2+} ATPase (Section 13.2). On binding IP_3, specific IP_3-gated Ca^{2+}–channel proteins in the ER membrane open to allow calcium ions to flow from the ER into the cytoplasm. Calcium ion is itself a signaling molecule: It can bind proteins, including a ubiquitous signaling protein called calmodulin and enzymes such as protein kinase C. By such means, the elevated level of cytoplasmic Ca^{2+} triggers processes such as smooth-muscle contraction, glycogen breakdown, and vesicle release.

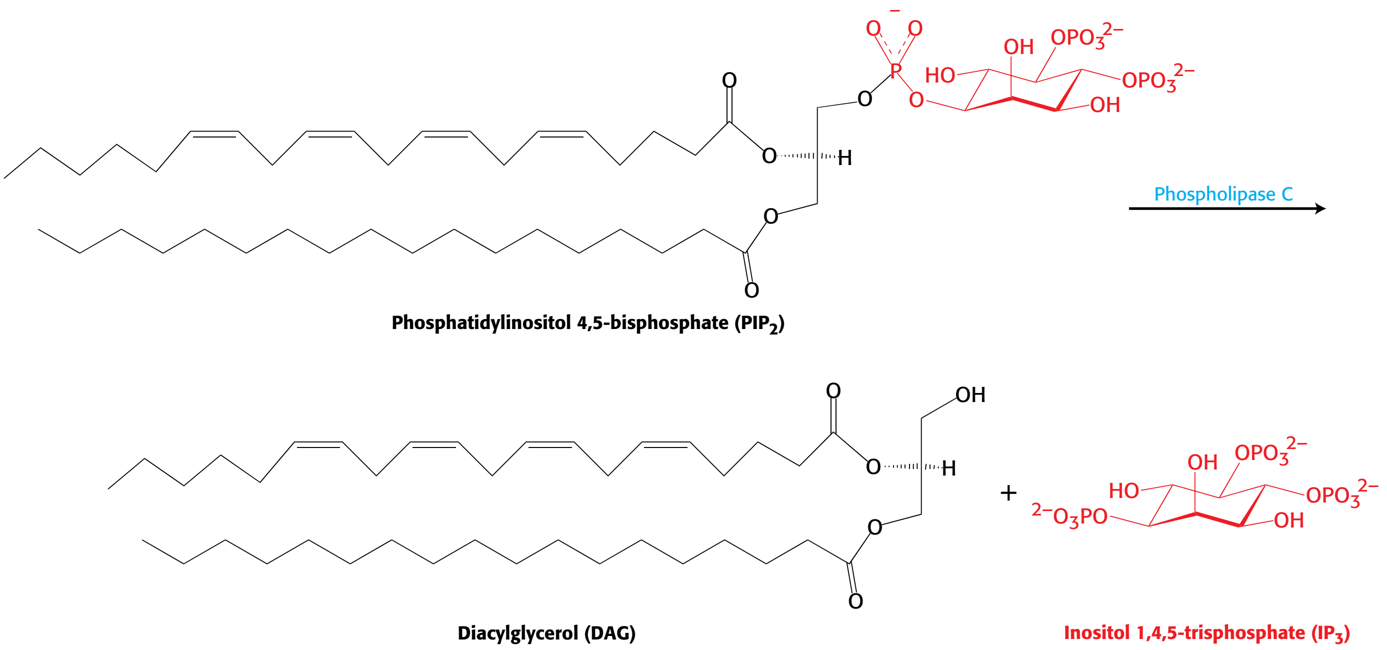

Phospholipase C
Phosphatidylinositol 4,5-bisphosphate (PIP$_2$)
Diacylglycerol (DAG)
+
Inositol 1,4,5-trisphosphate (IP$_3$)

FIGURE 14.13 Phospholipase C reaction. Phospholipase C cleaves the membrane lipid phosphatidylinositol 4,5-bisphosphate into two second messengers: diacylglycerol, which remains in the membrane, and inositol 1,4,5-trisphosphate, which diffuses away from the membrane.

DAG remains in the plasma membrane. There, it activates *protein kinase C* (PKC), a protein kinase that phosphorylates serine and threonine residues in many target proteins. To bind DAG, the specialized DAG-binding domains of this kinase require bound calcium. Note that DAG and IP$_3$ work in tandem: IP$_3$ increases the Ca^{2+} concentration, and Ca^{2+} facilitates the DAG-mediated activation of protein kinase C. The phosphoinositide cascade is summarized in Figure 14.14. Both IP$_3$ and DAG act transiently because they are converted into other species by phosphorylation or other processes.

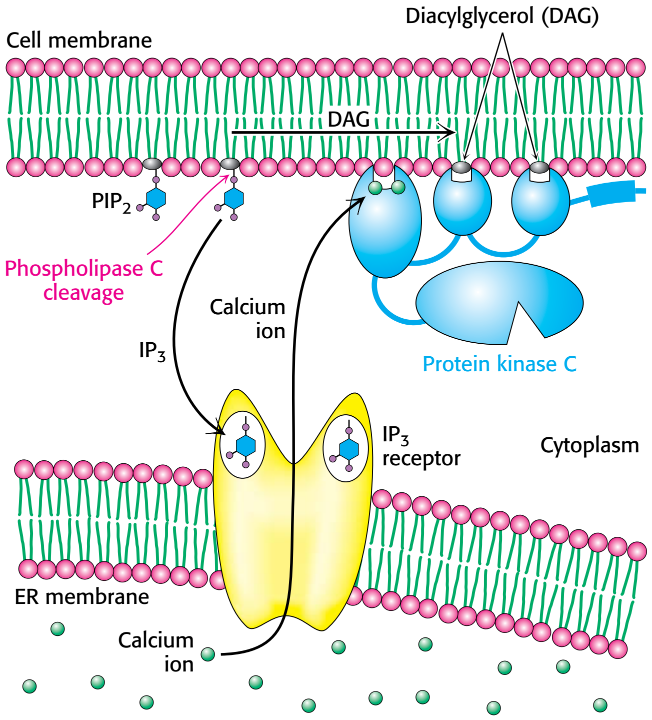


FIGURE 14.14 Phosphoinositide cascade. The cleavage of phosphatidylinositol 4,5-bisphosphate (PIP$_2$) into diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP$_3$) results in the release of calcium ions (owing to the opening of the IP$_3$ receptor ion channels) and the activation of protein kinase C (owing to the binding of protein kinase C to free DAG in the membrane). Calcium ions bind to protein kinase C and help facilitate its activation.

Calcium ion is a widely used second messenger

Calcium ion participates in many signaling processes in addition to the phosphoinositide cascade. Several properties of this ion account for its widespread use as an intracellular messenger. First, fleeting changes in Ca^{2+} concentration are readily detected. At steady state, intracellular levels of Ca^{2+} must be kept low to prevent the precipitation of carboxylated and phosphorylated compounds, which form poorly soluble salts with Ca^{2+}. Transport systems extrude Ca^{2+} from the cytoplasm, maintaining the cytoplasmic concentration of Ca^{2+} at approximately 100 nM, several orders of magnitude lower than that of the extracellular medium (Section 13.2). Given this low steady-state level, transient increases in Ca^{2+} concentration produced by signaling events can be readily sensed.

A second property of Ca^{2+} that makes it a highly suitable intracellular messenger is that it can bind tightly to proteins and induce substantial structural rearrangements. Calcium ions bind well to negatively charged oxygen atoms (from the side chains of glutamate and aspartate) and uncharged oxygen atoms (main-chain carbonyl

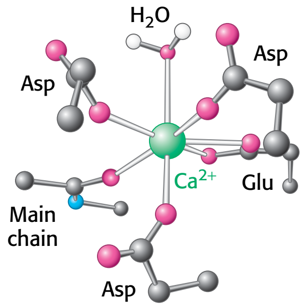
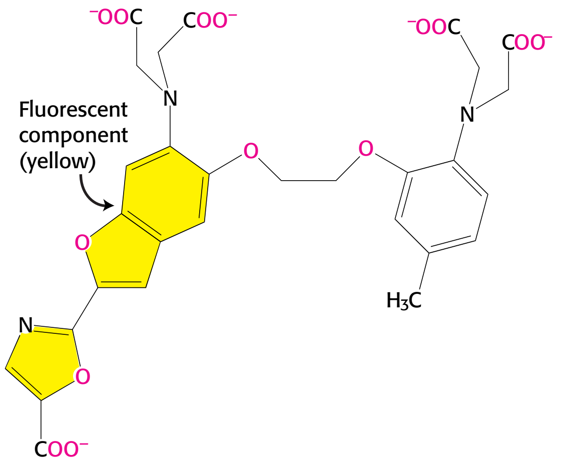


FIGURE 14.15 Calcium-binding site. In one common mode of binding, calcium is coordinated to six oxygen atoms of a protein and one (top) of water.

groups and side-chain oxygen atoms from glutamine and asparagine; Figure 14.15). *The capacity of Ca^{2+} to be coordinated to multiple ligands—from six to eight oxygen atoms—enables it to cross-link different segments of a protein and induce significant conformational changes.*

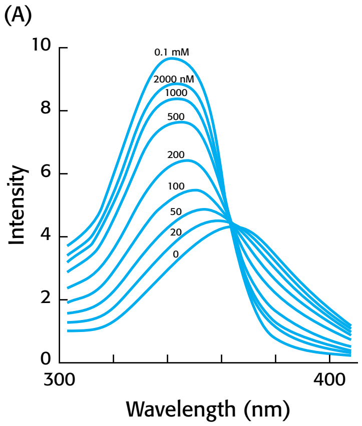
Our understanding of the role of Ca^{2+} in cellular processes has been greatly enhanced by our ability to detect changes in Ca^{2+} concentrations inside cells and even monitor these changes in real time. This ability depends on the use of specially designed dyes such as Fura-2 that bind Ca^{2+} and change their fluorescent properties on Ca^{2+} binding. Fura-2 binds Ca^{2+} through appropriately positioned oxygen atoms (shown in red) within its structure.

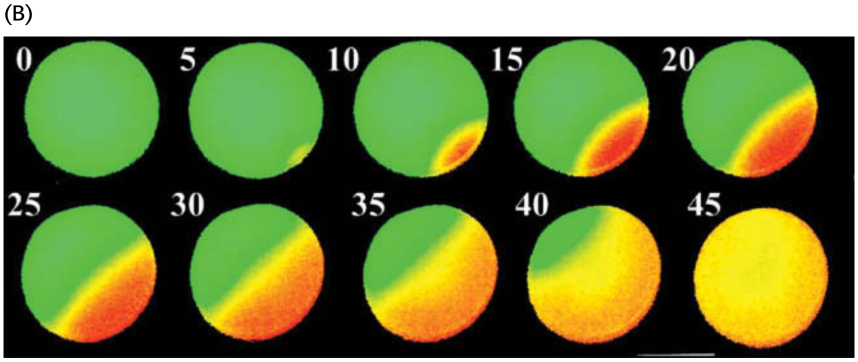


Fura-2

When such a dye is introduced into cells, changes in available Ca^{2+} concentration can be monitored with microscopes capable of detecting changes in fluorescence (Figure 14.16). Probes for sensing other second messengers such as cAMP also have been developed. These *molecular-imaging agents* are greatly enhancing our understanding of signal-transduction processes.

FIGURE 14.16 Calcium imaging. (A) The fluorescence spectra of the calcium-binding dye Fura-2 can be used to measure available calcium ion concentrations in solution and in cells. (B) A series of images show Ca^{2+} spreading across an egg cell following fertilization by sperm. These images were obtained through the use of Fura-2. The images are false colored: orange represents high Ca^{2+} concentrations, and green represents low Ca^{2+} concentrations. [(A) Information from S. J. Lippard and J. M. Berg, *Principles of Bioinorganic Chemistry* (University Science Books, 1994), p. 193; (B) data from Carroll M. et al., *J. Cell Sci.* 116: 4997–5004, 2003.]





Calcium ion often activates the regulatory protein calmodulin

Calmodulin (CaM), a 17-kDa protein with four Ca^{2+}-binding sites, serves as a calcium sensor in nearly all eukaryotic cells. *At cytoplasmic concentrations above about 500 nM, Ca^{2+} binds to and activates calmodulin.* Calmodulin is a member of the *EF-hand protein family*. The *EF hand* is a Ca^{2+}-binding motif that consists of a helix, a loop, and a second helix. This motif, originally discovered in the protein parvalbumin, was named the EF hand because the two key helices designated E and F in parvalbumin are positioned like the forefinger and thumb of the right hand (Figure 14.17). These two helices and the intervening loop form the Ca^{2+}-binding motif. Seven oxygen atoms are coordinated to each Ca^{2+}, six from the protein and one from a bound water molecule. Calmodulin is made up of four EF-hand motifs, each of which can bind a single Ca^{2+} ion.

The binding of Ca^{2+} to calmodulin induces substantial conformational changes in its EF hands, exposing hydrophobic surfaces that can be used to bind other proteins. Using its two sets of two EF hands, calmodulin clamps down around specific regions of target proteins, usually exposed α helices with appropriately positioned hydrophobic and charged groups (Figure 14.18). The Ca^{2+}–calmodulin complex stimulates a wide variety of enzymes, pumps, and other target proteins by inducing structural rearrangements in these binding partners. An especially noteworthy set of targets are several *calmodulin-dependent protein kinases* (CaM kinases) that phosphorylate many different proteins and regulate fuel metabolism, ionic permeability, neurotransmitter synthesis, and neurotransmitter release. We see here a recurring theme in signal-transduction pathways: the concentration of a second messenger is increased (in this case, Ca^{2+}); the signal is sensed by a second-messenger-binding protein (in this case, calmodulin); and the second-messenger-binding protein acts to generate changes in enzymes (in this case, calmodulin-dependent kinases) that control effectors.

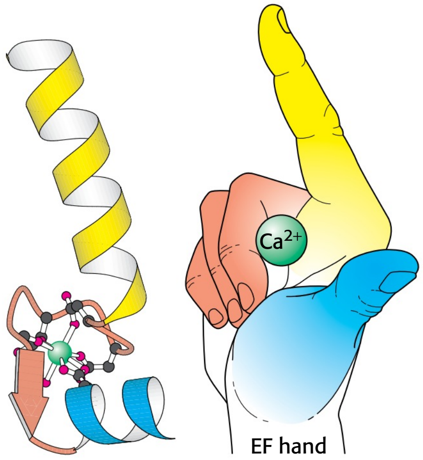


FIGURE 14.17 EF hand. Formed by a helix-loop-helix unit, an EF hand is a binding site for Ca^{2+} in many calcium-sensing proteins. Here, the E helix is yellow, the F helix is blue, and calcium is represented by the green sphere. *Note* that the calcium ion is bound in a loop connecting two nearly perpendicular helices. [Drawn from 1CLL.pdb.]

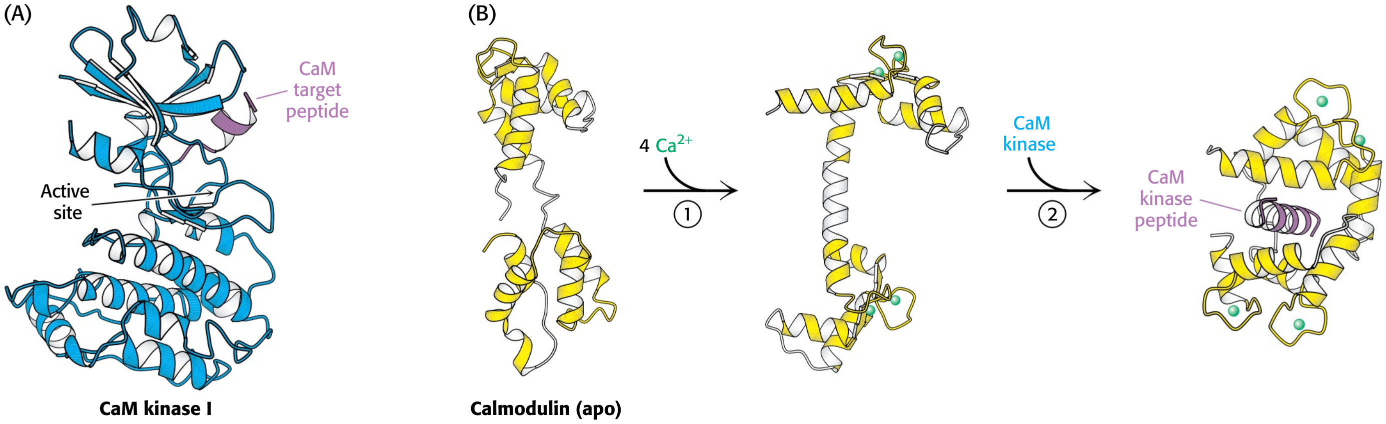


FIGURE 14.18 Calmodulin binds to α helices. (A) An α helix (purple) in CaM kinase I is a target for calmodulin. (B) On Ca^{2+} binding to the apo, or calcium-free, form of calmodulin (1), the two halves of calmodulin clamp down around the target helix (2), binding it through hydrophobic and ionic interactions. In CaM kinase I, this interaction allows the enzyme to adopt an active conformation. [Drawn from 1A06, 1CFD, 1CLL, and 1CM1.pdb.]

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes

The signaling pathways that we have examined so far have activated a protein kinase as a downstream component of the pathway. We now turn to a class of signal-transduction pathways that are *initiated by receptors that* *include protein kinases as part of their structures*. The activation of these protein kinases sets in motion other processes that ultimately modify the effectors of these pathways.

An example is the signal-transduction pathway initiated by *insulin,* the hormone released in response to increased blood-glucose levels after a meal. In all of its detail, this multifaceted pathway is quite complex. Hence, we will focus solely on the major branch, which leads to the mobilization of glucose transporters to the cell surface. These transporters allow the cell to take up the glucose that is plentiful in the blood stream after a meal.

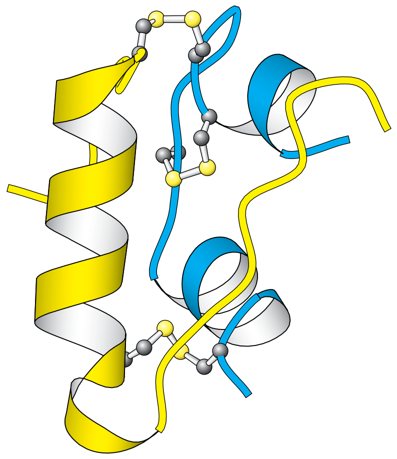


FIGURE 14.19 Insulin structure. *Notice* that insulin consists of two chains (shown in blue and yellow) linked by two interchain disulfide bonds. The α chain (blue) also has an intrachain disulfide bond. [Drawn from 1B2F.pdb.]

The insulin receptor is a dimer that closes around a bound insulin molecule

Insulin is a peptide hormone that consists of two chains that are linked by three disulfide bonds (Figure 14.19). Its receptor has a quite different structure from that of the β-AR. The *insulin receptor* is a dimer of two identical units. Each unit consists of one α chain and one β chain linked to one another by a single disulfide bond (Figure 14.20). Each α subunit lies completely outside the cell, whereas each β subunit lies primarily inside the cell, spanning the membrane with a single transmembrane segment. The two α subunits move together to form a binding site for a single insulin molecule, a surprising occurrence because two different surfaces on the insulin molecule must interact with the two identical insulin-receptor chains. The moving together of the dimeric units in the presence of an insulin molecule sets the signaling pathway in motion. *The closing up of an oligomeric receptor or the oligomerization of monomeric receptors around a bound ligand is a strategy used by many receptors to initiate a signal, particularly by those containing a protein kinase.*

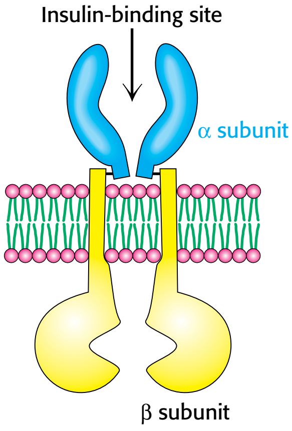
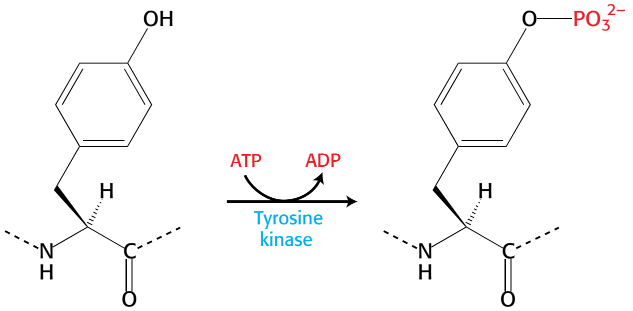


FIGURE 14.20 The insulin receptor. The receptor consists of two units, each of which consists of an α subunit and a β subunit linked by a disulfide bond. Two α subunits, which lie outside the cell, come together to form a binding site for insulin. Each β subunit lies primarily inside the cell and includes a protein kinase domain.

Each β subunit consists primarily of a protein kinase domain, homologous to protein kinase A. However, this kinase differs from protein kinase A in two important ways. First, the insulin-receptor kinase is a *tyrosine kinase;* that is, it catalyzes the transfer of a phosphoryl group from ATP to the hydroxyl group of tyrosine, rather than serine or threonine.



Because this tyrosine kinase is a component of the receptor itself, the insulin receptor is referred to as a *receptor tyrosine kinase*. Second, the insulin receptor kinase is in an inactive conformation when the domain is not covalently modified. The kinase is rendered inactive by the position of an unstructured loop (called the *activation loop*) that lies in the center of the structure.

Insulin binding results in the cross-phosphorylation and activation of the insulin receptor

When the two α subunits move together to surround an insulin molecule, the two protein kinase domains on the inside of the cell also are drawn together. It is important to note that as they come together, the flexible

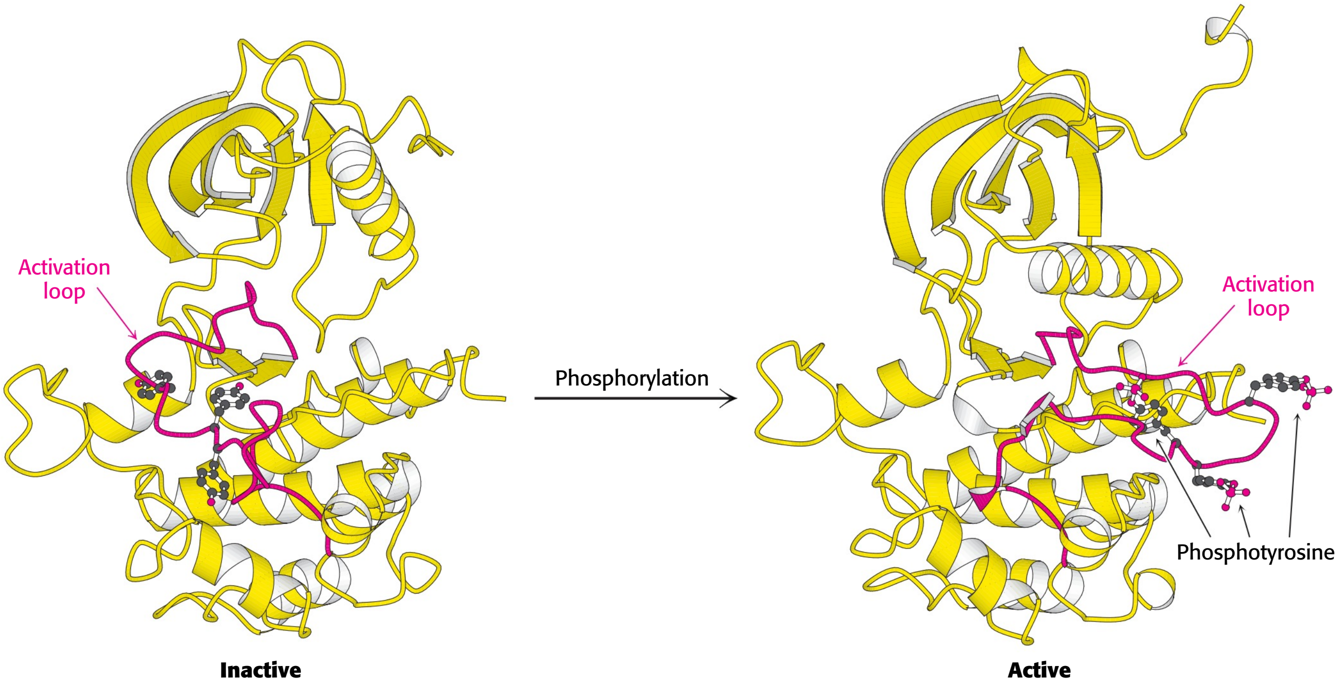


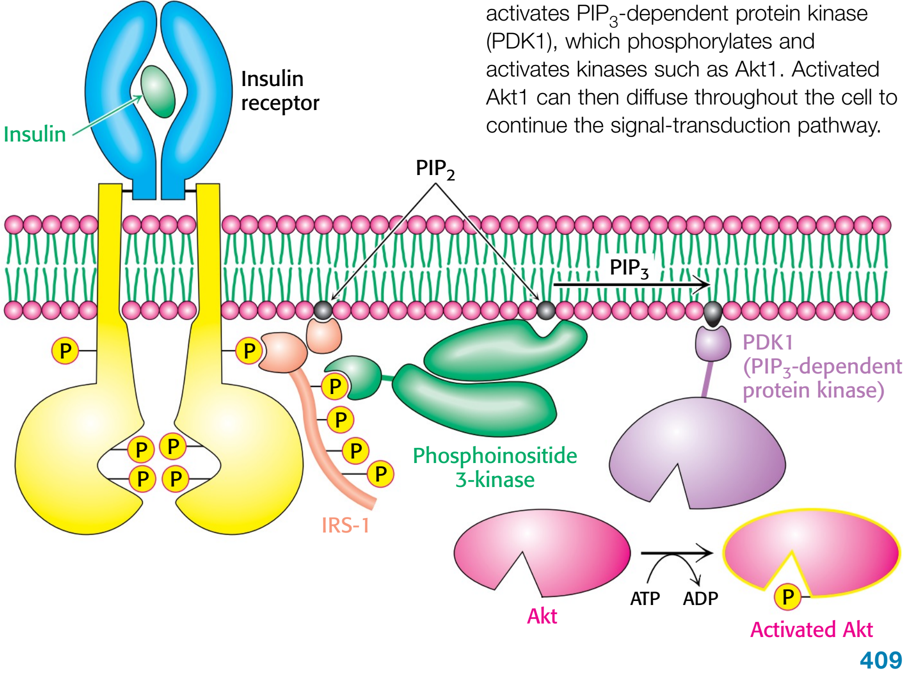
FIGURE 14.21 Activation of the insulin receptor by phosphorylation. The activation loop is shown in red in this model of the protein kinase domain of the β subunit of the insulin receptor. The unphosphorylated structure on the left is not catalytically active. *Notice* that, when three tyrosine residues in the activation loop are phosphorylated, the activation loop swings across the structure and the kinase structure adopts a more compact conformation. This conformation is catalytically active. [Drawn from 1IRK.pdb and 1IR3.pdb.]

activation loop of one kinase subunit is able to fit into the active site of the other kinase subunit within the dimer. With the two β subunits forced together, the kinase domains catalyze the addition of phosphoryl groups from ATP to tyrosine residues in the activation loops. When these tyrosine residues are phosphorylated, a striking conformational change takes place (Figure 14.21). The rearrangement of the activation loop converts the kinase into an active conformation. Thus, *insulin binding on the outside of the cell results in the activation of a membrane-associated kinase within the cell.*

The activated insulin-receptor kinase initiates a kinase cascade

On phosphorylation, the insulin-receptor tyrosine kinase is activated. Because the two units of the receptor are held in close proximity to one another, additional sites within the receptor also are phosphorylated. These phosphorylated sites act as docking sites for other substrates, including a class of molecules referred to as *insulin-receptor substrates* (IRS; Figure 14.22). IRS-1 and IRS-2 are two homologous

FIGURE 14.22 Insulin signaling. The binding of insulin results in the cross-phosphorylation and activation of the insulin receptor. Phosphorylated sites on the receptor act as binding sites for insulin-receptor substrates such as IRS-1. The lipid kinase phosphoinositide 3-kinase binds to phosphorylated sites on IRS-1 through its regulatory domain, then converts PIP_2 into PIP_3. Binding to PIP_3 activates PIP_3-dependent protein kinase (PDK1), which phosphorylates and activates kinases such as Akt1. Activated Akt1 can then diffuse throughout the cell to continue the signal-transduction pathway.



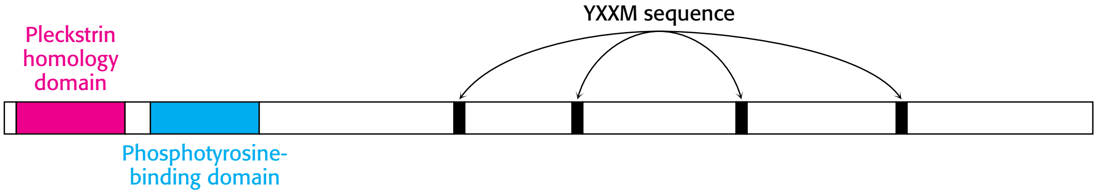


FIGURE 14.23 The modular structure of insulin-receptor substrates IRS-1 and IRS-2. This schematic view represents the amino acid sequence common to IRS-1 and IRS-2. Each protein contains a pleckstrin homology domain (which binds phosphoinositide lipids), a phosphotyrosine-binding domain, and four sequences that approximate Tyr-X-X-Met (YXXM). The four sequences are phosphorylated by the insulin-receptor tyrosine kinase.

proteins with a common modular structure (Figure 14.23). The amino-terminal part includes a *pleckstrin homology domain,* which binds phosphoinositide, and a phosphotyrosine-binding domain. These domains act together to anchor the IRS protein to the insulin receptor and the associated membrane. Each IRS protein contains four sequences that approximate the form Tyr-X-X-Met. These sequences are also substrates for the activated insulin-receptor kinase. When the tyrosine residues within these sequences are phosphorylated to become phosphotyrosine residues, IRS molecules can act as *adaptor proteins:* they are not enzymes but serve to tether the downstream components of this signaling pathway to the membrane.

Phosphotyrosine residues, such as those in the IRS proteins, are recognized most often by *Src homology 2* (SH2) domains (Figure 14.24). These domains, present in many signal-transduction proteins, bind to stretches of polypeptide that contain phosphotyrosine residues. Each specific SH2 domain shows a binding preference for phosphotyrosine in a particular sequence context. Which proteins contain SH2 domains that bind to phosphotyrosine-containing sequences in the IRS proteins? The most important of them are in a class of lipid kinases, called phosphoinositide 3-kinases (PI3Ks), that add a phosphoryl group to the 3-position of inositol in phosphatidylinositol 4,5-bisphosphate (PIP_2; Figure 14.25). These enzymes are heterooligomers that consist of 110-kDa catalytic subunits and 85-kDa regulatory subunits. Through SH2 domains in the regulatory subunits, these enzymes bind to the IRS proteins and are drawn to the membrane where they can phosphorylate PIP_2 to form phosphatidylinositol 3,4,5-trisphosphate (PIP_3). PIP_3, in turn, activates a protein kinase, PDK1, by virtue of a pleckstrin homology domain present in this kinase that is specific for PIP_3 (Figure 14.22). The activated PDK1 phosphorylates and activates Akt, another protein kinase. Akt is not membrane anchored and moves through the cell to phosphorylate targets that include components that control the trafficking of the glucose receptor GLUT4 to the cell surface as well as enzymes that stimulate glycogen synthesis (Section 21.4).

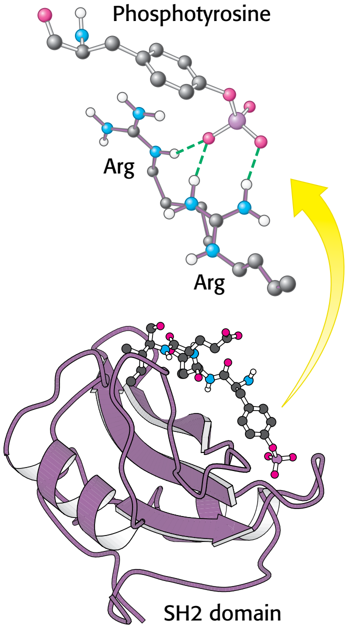


FIGURE 14.24 Structure of the SH2 domain. The domain is shown bound to a phosphotyrosine-containing peptide. *Notice* at the top that the negatively charged phosphotyrosine residue interacts with two arginine residues that are conserved in essentially all SH2 domains. [Drawn from 1SPS.pdb.]

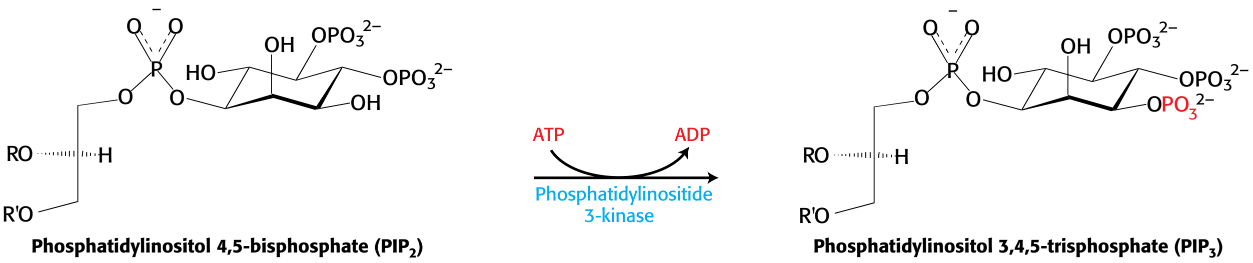

ATP
ADP
Phosphatidylinositide 3-kinase
Phosphatidylinositol 4,5-bisphosphate (PIP2)
Phosphatidylinositol 3,4,5-trisphosphate (PIP3)

FIGURE 14.25 Action of a lipid kinase in insulin signaling. Phosphorylated IRS-1 and IRS-2 activate the enzyme phosphatidylinositide 3-kinase, an enzyme that converts PIP_2 into PIP_3.

The cascade initiated by the binding of insulin to the insulin receptor is summarized in Figure 14.26. The signal is amplified at several stages along this pathway. Because the activated insulin receptor itself is a protein kinase, each activated receptor can phosphorylate multiple IRS molecules. Activated enzymes further amplify the signal in at least two of the subsequent steps. Thus, a small increase in the concentration of circulating insulin can produce a robust intracellular response. Note that although the insulin pathway described here may seem complicated, it is substantially less elaborate than the full signaling network initiated by insulin.

Insulin signaling is terminated by the action of phosphatases

We have seen that the activated G protein promotes its own inactivation by the release of a phosphoryl group from GTP. In contrast, proteins phosphorylated on serine, threonine, or tyrosine residues are extremely stable kinetically. Specific enzymes, called protein phosphatases, are required to hydrolyze these phosphorylated proteins and return them to their initial states. Similarly, lipid phosphatases are required to remove phosphoryl groups from inositol lipids that had been activated by lipid kinases. In insulin signaling, three classes of enzymes are of particular importance in shutting off the signaling pathway: (1) protein tyrosine phosphatases that remove phosphoryl groups from tyrosine residues on the insulin receptor and the IRS adaptor proteins, (2) lipid phosphatases that hydrolyze PIP_3 to PIP_2, and (3) protein serine phosphatases that remove phosphoryl groups from activated protein kinases such as Akt. Many of these phosphatases are activated or recruited as part of the response to insulin. Thus, the binding of the initial signal sets the stage for the eventual termination of the response.

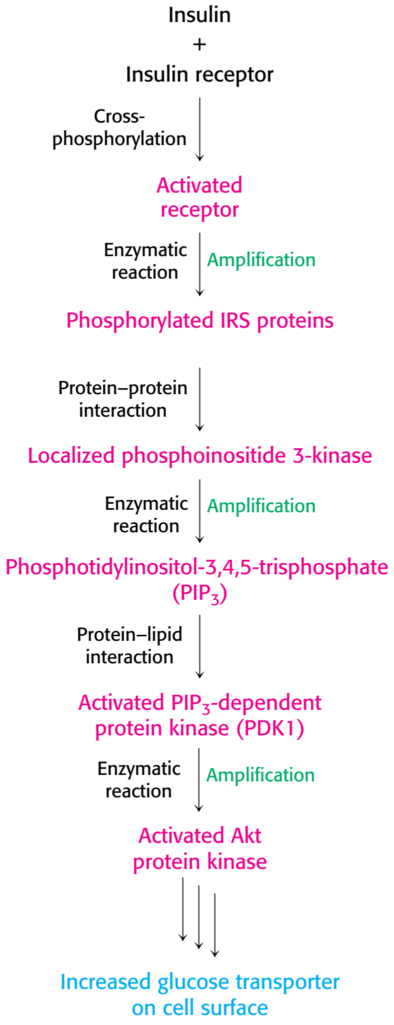

Insulin
+
Insulin receptor
Cross-phosphorylation
Activated receptor
Enzymatic reaction
Amplification
Phosphorylated IRS proteins
Protein–protein interaction
Localized phosphoinositide 3-kinase
Enzymatic reaction
Amplification
Phosphotidylinositol-3,4,5-trisphosphate (PIP_3)
Protein–lipid interaction
Activated PIP_3-dependent protein kinase (PDK1)
Enzymatic reaction
Amplification
Activated Akt protein kinase
Increased glucose transporter on cell surface

FIGURE 14.26 Insulin signaling pathway. Key steps in the signal-transduction pathway initiated by the binding of insulin to the insulin receptor.

14.3 EGF Signaling: Signal-Transduction Pathways Are Poised to Respond

Our consideration of the signal-transduction cascades initiated by epinephrine and insulin included examples of how components of signal-transduction pathways are poised for action, ready to be activated by minor modifications. For example, G-protein subunits require only the binding of GTP in exchange for GDP to transmit a signal. This exchange reaction is thermodynamically favorable, but it is quite slow in the absence of an appropriate activated 7TM receptor. Similarly, the tyrosine kinase domains of the dimeric insulin receptor are ready for phosphorylation and activation but require insulin bound between two α subunits to draw the activation loop of one tyrosine kinase into the active site of a partner tyrosine kinase to initiate the signaling cascade.

Next, we examine a signal-transduction pathway that reveals another clear example of how these signaling cascades are poised to respond. This pathway is activated by the signal molecule *epidermal growth factor* (EGF). Like that of the insulin receptor, the initiator of this pathway is a receptor tyrosine kinase. Both the extracellular and the intracellular domains of this receptor are ready for action, held in check only by a specific structure that prevents receptors from coming together. Furthermore, in the EGF pathway, we will encounter several additional classes of signaling components that participate in many other signaling networks.

EGF binding results in the dimerization of the EGF receptor

Epidermal growth factor is a 6-kDa polypeptide that stimulates the growth of epidermal and epithelial cells (Figure 14.27). The *EGF receptor* (EGFR),

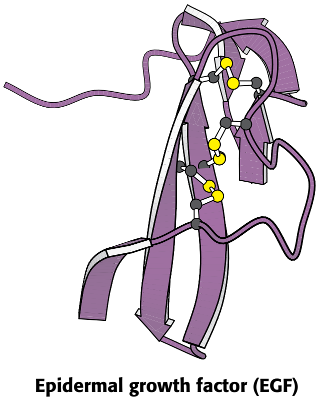

Epidermal growth factor (EGF)

FIGURE 14.27 Structure of epidermal growth factor. *Notice* that three intrachain disulfide bonds stabilize the compact three-dimensional structure of the growth factor. [Drawn from 1EGF.pdb.]

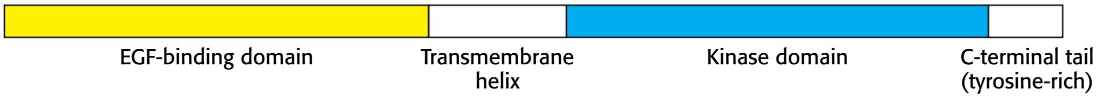


FIGURE 14.28 Modular structure of the EGF receptor. This schematic view of the amino acid sequence of the EGF receptor shows the EGF-binding domain that lies outside the cell, a single transmembrane helix-forming region, the intracellular tyrosine kinase domain, and the tyrosine-rich domain at the carboxyl terminus.

like the insulin receptor, is a dimer of two identical subunits. Each subunit contains an intracellular protein tyrosine kinase domain that participates in cross-phosphorylation reactions (Figure 14.28). Unlike those of the insulin receptor, however, these units exist as monomers until they bind EGF. Moreover, each EGF receptor monomer binds a single molecule of EGF in its extracellular domain (Figure 14.29). Thus the dimer binds two ligand molecules, in contrast with the insulin-receptor dimer, which binds only one ligand. Note that each EGF molecule lies far away from the dimer interface. This interface includes a so-called *dimerization arm* from each monomer that reaches out and inserts into a binding pocket on the other monomer.

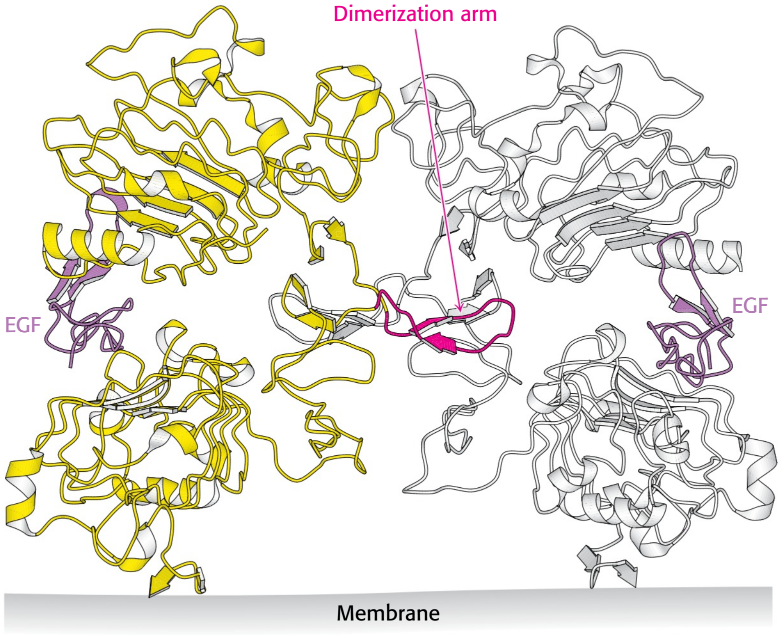


FIGURE 14.29 EGF receptor dimerization. The structure of the extracellular region of the EGF receptor is shown bound to EGF. *Notice* that the structure is dimeric with one EGF molecule bound to each receptor molecule and that the dimerization is mediated by a dimerization arm that extends from each receptor molecule. [Drawn from 1IVO.pdb.]

Although this structure nicely reveals the interactions that support the formation of a receptor dimer favoring cross-phosphorylation, it raises another question: Why doesn't the receptor dimerize and signal in the absence of EGF? This question has been addressed by examining the structure of the EGF receptor in the absence of bound ligand (Figure 14.30). This structure is, indeed, monomeric, and each monomer is in a conformation that is quite different from that observed in the ligand-bound dimer. In particular, the dimerization arm binds to a domain *within the same monomer* that holds the receptor in a closed configuration. In essence, the receptor is poised in a spring-loaded conformation held in position by the contact between the interaction loop and another part of the structure, ready to bind ligand and change into a conformation active for dimerization and signaling.

This observation suggests that a receptor that exists in the extended conformation even in the absence of bound ligand would be constitutively active. Remarkably, such a receptor exists. This receptor, HER2, is approximately 50% identical in amino acid sequence with the EGF receptor and has the same domain structure. HER2 does not bind any known ligand, yet crystallographic studies reveal that it adopts an extended structure very similar to that observed for the ligand-bound EGF receptor. Under normal conditions, HER2 forms heterodimers with the EGF receptor and other members of the EGF receptor family and participates in cross-phosphorylation reactions with these receptors. HER2 is overexpressed in some cancers, presumably contributing to tumor growth by forming homodimers that signal even in the absence of ligand. We will return to HER2 when we consider approaches to cancer treatment based on knowledge of signaling pathways (Section 14.5).

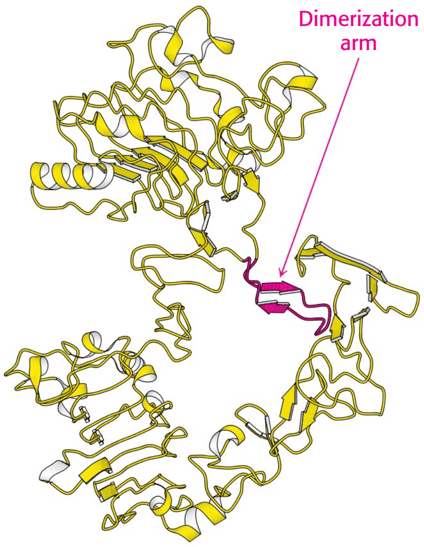


FIGURE 14.30 Structure of the unactivated EGF receptor. The extracellular domain of the EGF receptor is shown in the absence of bound EGF. *Notice* that the dimerization arm is bound to a part of the receptor that makes it unavailable for interaction with the other receptor. [Drawn from 1NQL.pdb.]

The EGF receptor undergoes phosphorylation of its carboxyl-terminal tail

Like the insulin receptor, the EGF receptor undergoes cross-phosphorylation of one unit by another unit within a dimer. However, unlike that of the insulin receptor, the site of this phosphorylation is not within the activation loop of the kinase, but rather in a region that lies on the C-terminal side of the kinase domain. As many as five tyrosine residues in this region are phosphorylated. The dimerization of the EGF receptor brings the C-terminal region on one receptor into the active site of its partner's kinase. The kinase itself is in an active conformation without phosphorylation, revealing again how this signaling system is poised to respond.

EGF signaling leads to the activation of Ras, a small G protein

The phosphotyrosines on the EGF receptors act as docking sites for SH2 domains on other proteins. The intracellular signaling cascade begins with the binding of *Grb2,* a key adaptor protein that contains one SH2 domain and two *Src homology 3* (SH3) domains. On phosphorylation of the receptor, the SH2 domain of Grb2 binds to the phosphotyrosine residues of the receptor tyrosine kinase. Through its two SH3 domains, Grb2 then binds polyproline-rich polypeptides within a protein called *Sos*. Sos, in turn, binds to *Ras* and activates it. A very prominent signal-transduction component, Ras is a member of a class of proteins called the *small G proteins.* Like the G proteins described in Section 14.1, the small G proteins contain bound GDP in their unactivated forms. Sos opens up the nucleotide-binding pocket of Ras, allowing GDP to escape and GTP to enter in its place. Because of its effect on Ras, Sos is referred to as a *guanine-nucleotide-exchange factor* (GEF). Thus, the binding of EGF to its receptor leads to the conversion of Ras into its GTP form through the intermediacy of Grb2 and Sos (Figure 14.31).

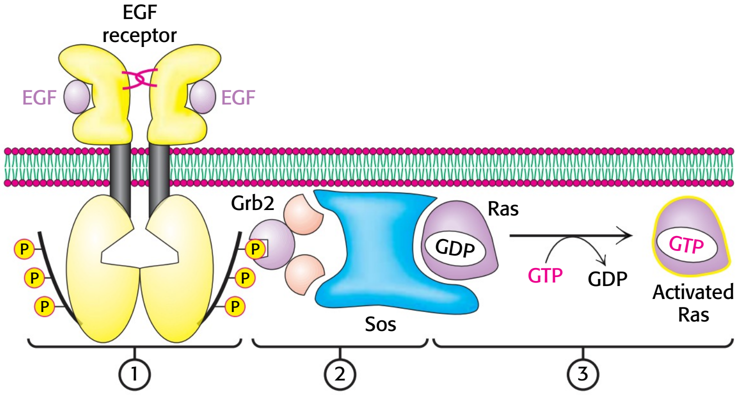


FIGURE 14.31 Ras activation mechanism. The dimerization of the EGF receptor due to EGF binding leads to: (1) the phosphorylation of the C-terminal tails of the receptor, (2) the subsequent recruitment of Grb2 and Sos, and (3) the exchange of GTP for GDP in Ras. This signal-transduction pathway results in the conversion of Ras into its activated GTP-bound form.

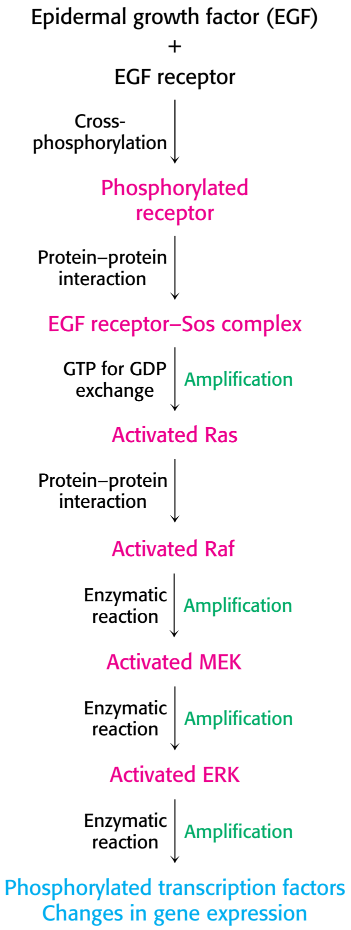

Epidermal growth factor (EGF)
+
EGF receptor
Cross-phosphorylation
Phosphorylated receptor
Protein–protein interaction
EGF receptor–Sos complex
GTP for GDP exchange
Amplification
Activated Ras
Protein–protein interaction
Activated Raf
Enzymatic reaction
Amplification
Activated MEK
Enzymatic reaction
Amplification
Activated ERK
Enzymatic reaction
Amplification
Phosphorylated transcription factors
Changes in gene expression

FIGURE 14.32 EGF signaling pathway. The key steps in the pathway initiated by EGF binding to the EGF receptor. A kinase cascade leads to the phosphorylation of transcription factors and concomitant changes in gene expression.

Activated Ras initiates a protein kinase cascade

Ras changes conformation when it is transformed from its GDP into its GTP form. In the GTP form, Ras binds other proteins, including a protein kinase termed *Raf*. When bound to Ras, Raf undergoes a conformational change that activates the Raf protein kinase domain. Both Ras and Raf are anchored to the membrane through covalently bound lipid modifications. Activated Raf then phosphorylates other proteins, including protein kinases termed MEKs. In turn, MEKs activate kinases called *extracellular signal-regulated kinases* (ERKs). ERKs then phosphorylate numerous substrates, including transcription factors in the nucleus as well as other protein kinases. The complete flow of information from the arrival of EGF at the cell surface to changes in gene expression is summarized in Figure 14.32.

Small G proteins, or small GTPases, constitute a large superfamily of proteins—grouped into subfamilies called Ras, Rho, Arf, Rab, and Ran—that play a major role in a host of cell functions including growth, differentiation, cell motility, cytokinesis (the separation of two cells during division), and the transport of materials throughout the cell (Table 14.2). As with the heterotrimeric G proteins, the small G proteins cycle between an active GTP-bound form and an inactive GDP-bound form. They differ from the heterotrimeric G proteins in being smaller (20–25 kDa versus 30–35 kDa) and monomeric. Nonetheless, the two families are related by divergent evolution, and small G proteins have many key mechanistic and structural motifs in common with the G_α subunit of the heterotrimeric G proteins.

EGF signaling is terminated by protein phosphatases and the intrinsic GTPase activity of Ras

Because so many components of the EGF signal-transduction pathway are activated by phosphorylation, we can expect protein phosphatases to play key roles in the termination of EGF signaling. Indeed, crucial phosphatases remove phosphoryl groups from tyrosine residues on the EGF receptor and from serine, threonine, and tyrosine residues in the protein kinases that participate in the signaling cascade. The signaling process itself sets in motion the events that activate many of these phosphatases. Consequently, signal activation also initiates signal termination.

Like the G proteins activated by 7TM receptors, Ras possesses intrinsic GTPase activity. Thus, the activated GTP form of Ras spontaneously converts into the inactive GDP form. The rate of conversion can be accelerated in the presence of *GTPase-activating proteins* (GAPs), which interact with small G proteins in the GTP form and facilitate GTP hydrolysis. Thus, the lifetime of activated Ras is regulated by accessory proteins in the cell. The GTPase activity of Ras is crucial for shutting off signals leading to cell growth, and so it is not surprising that mutations in Ras are found in many types of cancer, as discussed in Section 14.5.

TABLE 14.2 Ras superfamily of GTPases

Subfamily	Function
Ras	Regulates cell growth through serine–threonine protein kinases
Rho	Reorganizes cytoskeleton through serine–threonine protein kinases
Arf	Activates the ADP-ribosyltransferase of the cholera toxin A subunit; regulates vesicular trafficking pathways; activates phospholipase D
Rab	Plays a key role in secretory and endocytotic pathways
Ran	Functions in the transport of RNA and protein into and out of the nucleus

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways

We can begin to make sense of the complexity of signal-transduction pathways by taking note of several common themes that have appeared consistently in the pathways described in this chapter and underlie many additional signaling pathways not considered herein.

1. *Protein kinases are central to many signal-transduction pathways*. Protein kinases are central to all three signal-transduction pathways described in this chapter. In the epinephrine-initiated pathway, cAMP-dependent protein kinase (PKA) lies at the end of the pathway, transducing information represented by an increase in cAMP concentration into covalent modifications that alter the activity of key metabolic enzymes. In the insulin- and EGF-initiated pathways, the receptors themselves are protein kinases and several additional protein kinases participate downstream in the pathways. Signal amplification due to protein kinase cascades is a feature common to all three pathways. Although not presented in this chapter, protein kinases often phosphorylate multiple substrates and are thus able to generate a diversity of responses.

2. *Second messengers participate in many signal-transduction pathways.* We have encountered several second messengers, including cAMP, Ca^{2+}, IP_3, and the lipid DAG. Because second messengers are generated by enzymes or by the opening of ion channels, their concentrations can be tremendously amplified compared with the signals that lead to their generation. Specialized proteins sense the concentrations of these second messengers and continue the flow of information along signal-transduction pathways.

The second messengers that we have seen recur in many additional signal-transduction pathways. For example, in a consideration of the sensory systems in Chapter 33, we will see how Ca^{2+}-based signaling and cyclic nucleotide-based signaling play key roles in vision and olfaction.

3. *Specialized domains that mediate specific interactions are present in many signaling proteins.* The "wiring" of many signal-transduction pathways is based on particular protein domains that mediate the interactions between protein components of a particular signaling cascade. We have encountered several of them, including pleckstrin homology domains, which facilitate protein interactions with the lipid PIP_3; SH2 domains, which mediate interactions with polypeptides containing phosphorylated tyrosine residues; and SH3 domains, which interact with peptide sequences that contain multiple proline residues. Many other such domain families exist. In many cases, individual members of each domain family have unique features that allow them to bind to their targets only within a particular sequence context, making them specific for a given signaling pathway and avoiding unwanted cross-talk. *Signal-transduction pathways have evolved in large part by the incorporation of DNA fragments encoding these domains into genes encoding pathway components.*

The presence of these domains is tremendously helpful to scientists trying to unravel signal-transduction pathways. When a protein in a signal-transduction pathway is identified, its amino acid sequence can be analyzed for the presence of these specialized domains by the methods described in Chapter 6. If one or more domains of known function is found, it is often possible to develop clear hypotheses about potential binding partners and signal-transduction mechanisms.

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases

In light of their complexity, it comes as no surprise that signal-transduction pathways occasionally fail, leading to disease states. Cancer, a set of diseases characterized by uncontrolled or inappropriate cell growth, is strongly associated with defects in signal-transduction proteins. Indeed, the study of cancer, particularly cancers caused by certain viruses, has contributed greatly to our understanding of signal-transduction proteins and pathways.

For example, Rous sarcoma virus is a retrovirus that causes sarcoma (a cancer of tissues of mesodermal origin such as muscle or connective tissue) in chickens. In addition to the genes necessary for viral replication, this virus carries a gene termed v-*src*. The v-*src* gene is an *oncogene;* it leads to the generation of cancerlike characteristics in susceptible cell types. The protein encoded by the v-*src* gene, v-Src, is a protein tyrosine kinase that includes SH2 and SH3 domains. The v-Src protein is similar in amino acid sequence to a protein normally found in chicken-muscle cells referred to as c-Src (for cellular Src; Figure 14.33A). The c-*src* gene does not induce cell transformation and is termed a *proto-oncogene,* referring to the fact that this gene, when mutated, can be converted into an oncogene. The protein that it encodes is a signal-transduction protein that regulates cell growth.

Why is the biological activity of the v-Src protein so different from that of c-Src? c-Src contains a key tyrosine residue near its C-terminal end that, when phosphorylated, is bound intramolecularly by the upstream SH2 domain (Figure 14.33B). This interaction maintains the kinase domain in an inactive conformation. However, in v-Src, the C-terminal 19 amino acids of c-Src are replaced by a completely different stretch of 11 amino acids that lacks this critical tyrosine residue. Thus, v-Src is always active and can promote unregulated cell growth. Since the discovery of Src, many other mutated protein kinases have been identified as oncogenes.

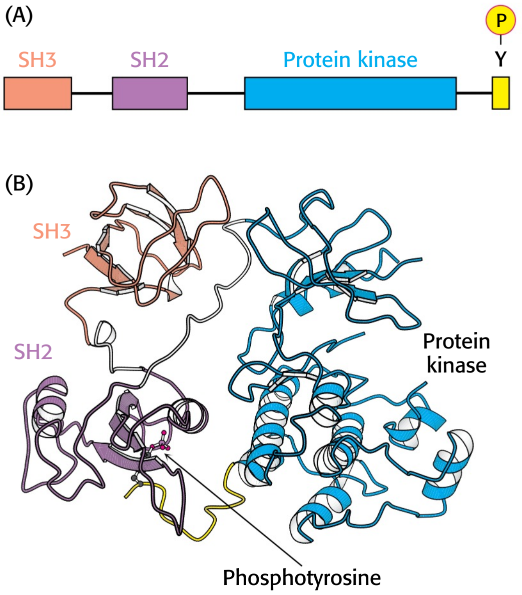


FIGURE 14.33 Src structure. (A) Cellular Src includes an SH3 domain, an SH2 domain, a protein kinase domain, and a carboxyl-terminal tail that includes a key tyrosine residue. (B) Structure of c-Src in an inactivated form with the key tyrosine residue phosphorylated. *Notice* how the three domains work together to keep the enzyme in an inactive conformation: the phosphotyrosine residue is bound in the SH2 domain and the linker between the SH2 domain and the protein kinase domain is bound by the SH3 domain. [Drawn from 2PTK.pdb.]

The gene encoding Ras, a component of the EGF-initiated pathway, is one of the genes most commonly mutated in human tumors. Mammalian cells contain three 21-kDa Ras proteins (H-, K-, and N-Ras), each of which cycles between inactive GDP and active GTP forms. The most common mutations in tumors lead to a loss of the ability to hydrolyze GTP. Thus, the Ras protein is trapped in the "on" position and continues to stimulate cell growth, even in the absence of a continuing signal.

Other genes can contribute to cancer development only when both copies of the gene normally present in a cell are deleted or otherwise damaged. Such genes are called *tumor-suppressor genes*. For example, genes for some of the phosphatases that participate in the termination of EGF signaling are tumor suppressors. Without any functional phosphatase present, EGF signaling persists once initiated, stimulating inappropriate cell growth.

Monoclonal antibodies can be used to inhibit signal-transduction pathways activated in tumors

Mutated or overexpressed receptor tyrosine kinases are frequently observed in tumors. For instance, the epidermal-growth-factor receptor (EGFR) is overexpressed in some human epithelial cancers, including breast, ovarian, and colorectal cancer.

Because some small amount of the receptor can dimerize and activate the signaling pathway even without binding to EGF, overexpression of the receptor increases the likelihood that a "grow and divide" signal will be inappropriately sent to the cell. This understanding of cancer-related signal-transduction pathways has led to a therapeutic approach that targets the EGFR. The strategy is to produce monoclonal antibodies to the extracellular domains of the offending receptors. One such antibody, cetuximab (Erbitux), has effectively targeted the EGFR in colorectal cancers. Cetuximab inhibits the EGFR by competing with EGF for the binding site on the receptor. Because the antibody sterically blocks the change in conformation that exposes the dimerization arm, the antibody itself cannot induce dimerization. The result is that the EGFR-controlled pathway is not initiated.

Cetuximab is not the only monoclonal antibody that has been developed to target a receptor tyrosine kinase. Trastuzumab (Herceptin) inhibits another EGFR family member, HER2, that is overexpressed in approximately 30% of breast cancers. Recall that this protein can signal even in the absence of ligand, so it is especially likely that overexpression will stimulate cell proliferation. Breast-cancer patients are now being screened for HER2 overexpression and treated with Herceptin as appropriate. Thus, this cancer treatment is tailored to the genetic characteristics of the tumor.

Protein kinase inhibitors can be effective anticancer drugs

The widespread occurrence of overactive protein kinases in cancer cells suggests that molecules that inhibit these enzymes might act as antitumor agents. For example, more than 90% of patients with chronic myelogenous leukemia (CML) show a specific chromosomal defect in cancer cells (Figure 14.34). The translocation of genetic material between chromosomes 9 and 22 causes the c-*abl* gene, which encodes a tyrosine kinase of the Src family, to be inserted into the *bcr* gene on chromosome 22. The result is the production of a fusion protein called Bcr-Abl that consists primarily of sequences for the c-Abl kinase. However, the *bcr-abl* gene is not regulated appropriately; it is expressed at higher levels than that of the gene encoding the normal c-Abl kinase, stimulating a growth-promoting pathway. Because of this overexpression, leukemia cells express a unique target for chemotherapy. A specific inhibitor of the Bcr-Abl kinase, Gleevec (STI-571, imatinib mesylate), has proved to be a highly effective treatment for patients suffering from CML. This approach to cancer chemotherapy is fundamentally distinct from most approaches, which target all rapidly growing cells, including normal ones. Because Gleevec targets tumor cells specifically, side effects caused by the impairment of normal dividing cells can be minimized. *Thus, our understanding of signal-transduction pathways is leading to conceptually new disease treatment strategies.*

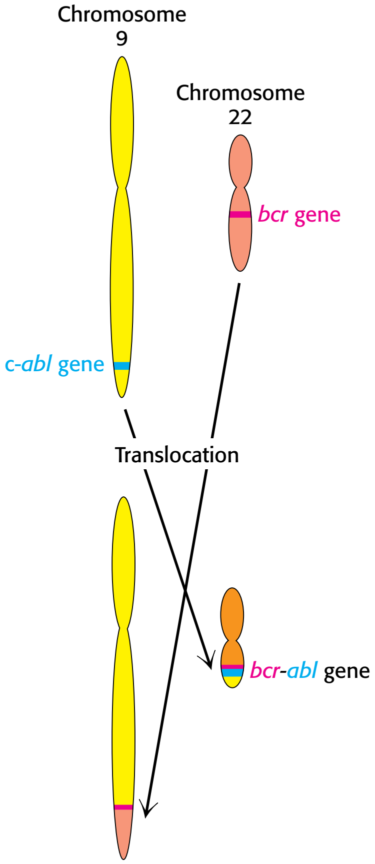


FIGURE 14.34 Formation of the *bcr-abl* gene by translocation. In chronic myelogenous leukemia, parts of chromosomes 9 and 22 are reciprocally exchanged, causing the *bcr* and *abl* genes to fuse. The protein kinase encoded by the *bcr-abl* gene is expressed at higher levels in tumor cells than is the c-*abl* gene in normal cells.

Cholera and whooping cough are the result of altered G-protein activity

Although defects in signal-transduction pathways have been most extensively studied in the context of cancer, such defects are important in many other diseases. Cholera and whooping cough are two pathologies of the G-protein-dependent signal pathways. Let us first consider the mechanism of action of the cholera toxin, secreted by the intestinal bacterium *Vibrio cholerae*. Cholera is a potentially life-threatening, acute diarrheal disease transmitted through contaminated water and food. It causes the voluminous secretion of electrolytes and fluids from the intestines of infected persons. The cholera toxin, also called *choleragen,* is a protein composed of two functional units—a β subunit that binds to G_{M1} gangliosides (Section 26.1) of the intestinal epithelium and a catalytic A subunit that enters the cell.

The A subunit catalyzes the covalent modification of a $G_{\alpha s}$ protein: the α subunit is modified by the attachment of an ADP-ribose to an arginine residue. This modification stabilizes the GTP-bound form of $G_{\alpha s}$, trapping the molecule in its active conformation. The active G protein, in turn, continuously activates protein kinase A. PKA opens a chloride channel and inhibits sodium absorption by the Na^+–H^+ exchanger by phosphorylating both the channel and the exchanger. The net result of the phosphorylation is an excessive loss of NaCl and the loss of large amounts of water into the intestine. Patients suffering from cholera may pass as much as twice their body weight in fluid in 4 to 6 days. Treatment consists of rehydration with a glucose–electrolyte solution.

Whereas cholera is a result of a G protein trapped in the active conformation, causing the signal-transduction pathway to be persistently stimulated, pertussis is a result of the opposite situation. Pertussis toxin is secreted by *Bordetella pertussis,* the bacterium responsible for whooping cough. Like choleragen, pertussis toxin adds an ADP-ribose moiety to a G_{α} subunit. However, in this case, the ADP-ribose group is added to a $G_{\alpha i}$ protein, a G_{α} subunit that inhibits adenylate cyclase, closes Ca^{2+} channels, and opens K^+ channels. The effect of this modification is to prevent binding of the heterotrimeric G_i protein to its receptor, trapping it in the "off" conformation. The pulmonary symptoms have not yet been traced to a particular target of the $G_{\alpha i}$ protein.

SUMMARY

In humans and other multicellular organisms, specific signal molecules are released from cells in one organ and are sensed by cells in other organs throughout the body. The message initiated by an extracellular ligand is converted into specific changes in metabolism or gene expression by means of often complex networks referred to as signal-transduction pathways. These pathways amplify the initial signal and lead to changes in the properties of specific effector molecules.

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves

Epinephrine binds to a cell-surface protein called the β-adrenergic receptor. This receptor is a member of the seven-transmembrane-helix receptor family, so named because each receptor has seven α helices that span the cell membrane. When epinephrine binds to the β-adrenergic receptor on the outside of the cell, the receptor undergoes a conformational change that is sensed inside the cell by a signaling protein termed a heterotrimeric G protein. The α subunit of the G protein exchanges a bound GDP molecule for GTP and concomitantly releases the heterodimer consisting of the β and γ subunits. The α subunit in the GTP form then binds to adenylate cyclase and activates it, leading to an increase in the concentration of the second messenger cyclic AMP. This increase in cyclic AMP concentration, in turn, activates protein kinase A. Other 7TM receptors also signal through heterotrimeric G proteins, although these pathways often include enzymes other than adenylate cyclase. One prominent pathway, the phosphoinositide pathway, leads to the activation of phospholipase C, which cleaves a membrane lipid to produce two secondary messengers, diacylglycerol and inositol 1,4,5-trisphosphate. An increased IP_3 concentration leads to the release of calcium ion, another important second messenger, into the cell. G-protein signaling is terminated by the hydrolysis of the bound GTP to GDP.

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes

Protein kinases are key components in many signal-transduction pathways, including some for which the protein kinase is an integral component of the initial receptor. An example of such a receptor is the membrane tyrosine kinase bound by insulin. Insulin binding causes one subunit within the dimeric receptor to phosphorylate specific tyrosine residues in the other subunit. The resulting conformational changes dramatically increase the kinase activity of the receptor. The activated receptor kinase initiates a signaling cascade that includes both lipid kinases and protein kinases. This cascade eventually leads to the mobilization of glucose transporters to the cell surface, increasing glucose uptake. Insulin signaling is terminated through the action of phosphatases.

14.3 EGF Signaling: Signal-Transduction Systems Are Poised to Respond

Only minor modifications are necessary to transform many signal-transduction proteins from their inactive into their active forms. Epidermal growth factor also signals through a receptor tyrosine kinase. EGF binding induces a conformational change that allows receptor dimerization and cross-phosphorylation. The phosphorylated receptor binds adaptor proteins that mediate the activation of Ras, a small G protein. Activated Ras initiates a protein kinase cascade that eventually leads to the phosphorylation of transcription factors and changes in gene expression. EGF signaling is terminated by the action of phosphatases and the hydrolysis of GTP by Ras.

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways

Protein kinases are components of many signal-transduction pathways, both as components of receptors and in other roles. Second messengers, including cyclic nucleotides, calcium, and lipid derivatives, are common in many signaling pathways. The changes in the concentrations of second messengers are often much larger than the changes associated with the initial signal owing to amplification along the pathway. Small domains that recognize phosphotyrosine residues or specific lipids are present in many signaling proteins and are essential to determining the specificity of interactions.

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases

Genes encoding components of signal-transduction pathways that control cell growth are often mutated in cancer. Some genes can be mutated to forms called oncogenes that are active regardless of appropriate signals. Monoclonal antibodies directed against cell-surface receptors that participate in signaling have been developed for use in cancer treatment. Our understanding of the molecular basis of cancer is leading to the development of anticancer drugs directed against specific targets, such as the specific kinase inhibitor Gleevec.

KEY TERMS

primary messenger (p. 398)
ligand (p. 398)
second messenger (p. 398)
cross talk (p. 398)
β-adrenergic receptor (β-AR) (p. 399)
seven-transmembrane-helix (7TM) receptor (p. 399)
rhodopsin (p. 400)
G protein (p. 400)
agonist (p. 401)
G-protein-coupled receptor (GPCR) (p. 402)

adenylate cyclase (p. 402)

protein kinase A (PKA) (p. 403)

β-adrenergic receptor kinase (p. 404)

phosphoinositide cascade (p. 404)

phosphatidylinositol 4,5-bisphosphate (PIP$_2$) (p. 404)

phospholipase C (p. 404)

protein kinase C (PKC) (p. 405)

calmodulin (CaM) (p. 407)

EF hand (p. 407)

calmodulin-dependent protein kinase (CaM kinase) (p. 407)

insulin (p. 408)

insulin receptor (p. 408)

tyrosine kinase (p. 408)

receptor tyrosine kinase (p. 408)

insulin-receptor substrate (IRS) (p. 409)

pleckstrin homology domain (p. 410)

adaptor protein (p. 410)

Src homology 2 (SH2) domain (p. 410)

epidermal growth factor (EGF) (p. 411)

EGF receptor (EGFR) (p. 411)

dimerization arm (p. 412)

Src homology 3 (SH3) domain (p. 413)

Ras (p. 413)

small G protein (p. 413)

guanine-nucleotide-exchange factor (GEF) (p. 413)

extracellular signal-regulated kinase (ERK) (p. 414)

GTPase-activating protein (GAP) (p. 414)

oncogene (p. 416)

proto-oncogene (p. 416)

tumor-suppressor gene (p. 416)

PROBLEMS

1. *Active mutants.* Some protein kinases are inactive unless they are phosphorylated on key serine or threonine residues. In some cases, active enzymes can be generated by mutating these serine or threonine residues to aspartate. Explain.

2. *In the pocket.* SH2 domains bind phosphotyrosine residues in deep pockets on their surfaces. Would you expect SH2 domains to bind phosphoserine and phosphothreonine with high affinity? Why or why not?

3. *On–off.* Why is the GTPase activity of G proteins crucial to the proper functioning of a cell? Why have G proteins not evolved to catalyze GTP hydrolysis more efficiently?

4. *Viva la différence.* Why is the fact that a monomeric hormone binds simultaneously to two identical receptor molecules, thus promoting the formation of a dimer of the receptor, considered remarkable?

5. *Antibodies mimicking hormones.* Antibodies have two identical antigen-binding sites. Remarkably, antibodies to the extracellular parts of growth-factor receptors often lead to the same cellular effects as does exposure to growth factors. Explain this observation.

6. *Facile exchange.* A mutated form of the α subunit of the heterotrimeric G protein has been identified; this form readily exchanges nucleotides even in the absence of an activated receptor. What would be the effect on a signaling pathway containing the mutated α subunit?

7. *Making connections.* Suppose that you were investigating a newly discovered growth-factor signal-transduction pathway. You found that, if you added GTPγS, a nonhydrolyzable analog of GTP, the duration of the hormonal response increased. What can you conclude?

8. *Diffusion rates.* Usually, rates of diffusion vary inversely with molecular weights; so smaller molecules diffuse faster than do larger ones. In cells, however, calcium ion diffuses more slowly than does cAMP. Propose a possible explanation.

9. *Negativity abounds.* Fura-2 is not effective for the study of calcium levels in intact, living cells. On the basis of how Fura-2 is depicted on p. 406, why is it ineffective?

10. *Awash with glucose.* Glucose is mobilized for ATP generation in muscle in response to epinephrine, which activates $G_{\alpha s}$. Cyclic AMP phosphodiesterase is an enzyme that converts cAMP into AMP. How would inhibitors of cAMP phosphodiesterase affect glucose mobilization in muscle?

11. *Getting it started.* The insulin receptor, on dimerization, cross-phosphorylates the activation loop of the other receptor molecule, leading to activation of the kinase. Propose how this phosphorylation event can take place if the kinase starts in an inactive conformation.

12. *Many defects.* Considerable effort has been directed toward determining the genes in which sequence variation contributes to the development of type 2 diabetes. Approximately 800 genes have been implicated. Propose an explanation for this observation.

13. *Growth-factor signaling.* Human growth hormone binds to a cell-surface membrane protein that is not a receptor tyrosine kinase. The intracellular domain of the receptor can bind other proteins inside the cell. Furthermore, studies indicate that the receptor is monomeric in the absence of hormone but dimerizes on hormone binding. Propose a possible mechanism for growth-hormone signaling.

14. *Receptor truncation.* You prepare a cell line that overexpresses a mutant form of EGFR in which the entire intracellular region of the receptor has been deleted. Predict the effect of overexpression of this construct on EGF signaling in this cell line.

15. *Hybrid.* Suppose that, through genetic manipulations, a chimeric receptor is produced that consists of the extracellular domain of the insulin receptor and the transmembrane and intracellular domains of the EGF receptor. Cells expressing this receptor are exposed to insulin, and the level of phosphorylation of the chimeric receptor is examined. What would you expect to observe and why? What would you expect to observe if these cells were exposed to EGF?

16. *Total amplification.* Suppose that each β-adrenergic receptor bound to epinephrine converts 100 molecules of $G_{\alpha s}$ into their GTP forms and that each molecule of activated adenylate cyclase produces 1000 molecules of cAMP per second. With the assumption of a full response, how many molecules of cAMP will be produced in 1 s after the formation of a single complex between epinephrine and the β-adrenergic receptor?

Chapter Integration Problems

17. *Nerve-growth-factor pathway.* Nerve-growth factor (NGF) binds to a protein tyrosine kinase receptor. The amount of diacylglycerol in the plasma membrane increases in cells expressing this receptor when treated with NGF. Propose a simple signaling pathway and identify the isoform of any participating enzymes. Would you expect the concentrations of any other common second messengers to increase on NGF treatment?

18. *Redundancy.* Because of the high degree of genetic variability in tumors, typically no single anticancer therapy is universally effective for all patients, even within a given tumor type. Hence, it is often desirable to inhibit a particular pathway at more than one point in the signaling cascade. In addition to the EGFR-directed monoclonal antibody cetuximab, propose alternative strategies for targeting the EGF signaling pathway for antitumor drug development.

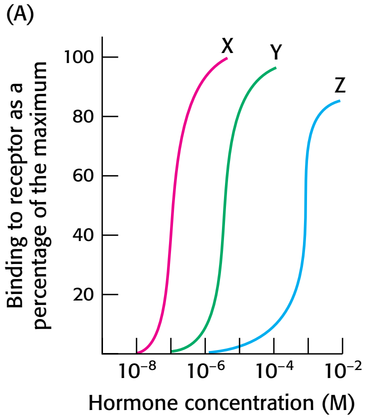
Mechanism Problems

19. *Distant relatives.* The structure of adenylate cyclase is similar to the structures of some types of DNA polymerases, suggesting that these enzymes derived from a common ancestor. Compare the reactions catalyzed by these two enzymes. In what ways are they similar?

20. *Kinase inhibitors as drugs.* Functional and structural analysis indicates that Gleevec is an ATP-competitive inhibitor of the Bcr-Abl kinase. In fact, many kinase inhibitors under investigation or currently marketed as drugs are ATP competitive. Can you suggest a potential drawback of drugs that utilize this particular mechanism of action?

Data Interpretation Problems

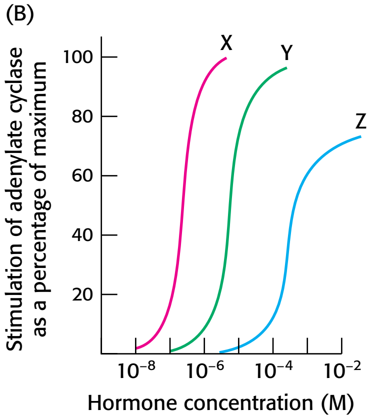
21. *Establishing specificity.* You wish to determine the hormone-binding specificity of a newly identified membrane receptor. Three different hormones, X, Y, and Z, were mixed with the receptor in separate experiments, and the percentage of binding capacity of the receptor was determined as a function of hormone concentration, as shown in graph A.



(a) What concentrations of each hormone yield 50% maximal binding?

(b) Which hormone shows the highest binding affinity for the receptor?

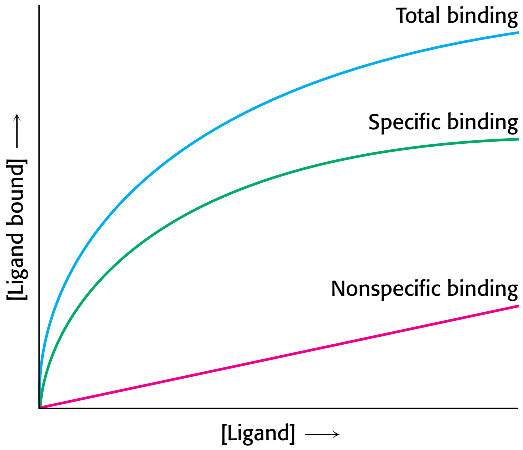
You next wish to determine whether the hormone–receptor complex stimulates the adenylate cyclase cascade. To do so, you measure adenylate cyclase activity as a function of hormone concentration, as shown in graph B.



(c) What is the relation between the binding affinity of the hormone–receptor complex and the ability of the hormone to enhance adenylate cyclase activity? What can you conclude about the mechanism of action of the hormone–receptor complex?

(d) Suggest experiments that would determine whether a $G_{\alpha s}$ protein is a component of the signal-transduction pathway.

22. *Binding issues.* A scientist wishes to determine the number of receptors specific for a ligand X, which he has in both radioactive and nonradioactive form. In one experiment, he adds increasing amounts of radioactive X and measures how much of it is bound to the cells. The result is shown as total activity in the following graph. Next, he performs the same experiment, except that he includes a several hundredfold excess of nonradioactive X. This result is shown as nonspecific binding. The difference between the two curves is the specific binding.



(a) Why is the total binding not an accurate representation of the number of receptors on the cell surface?

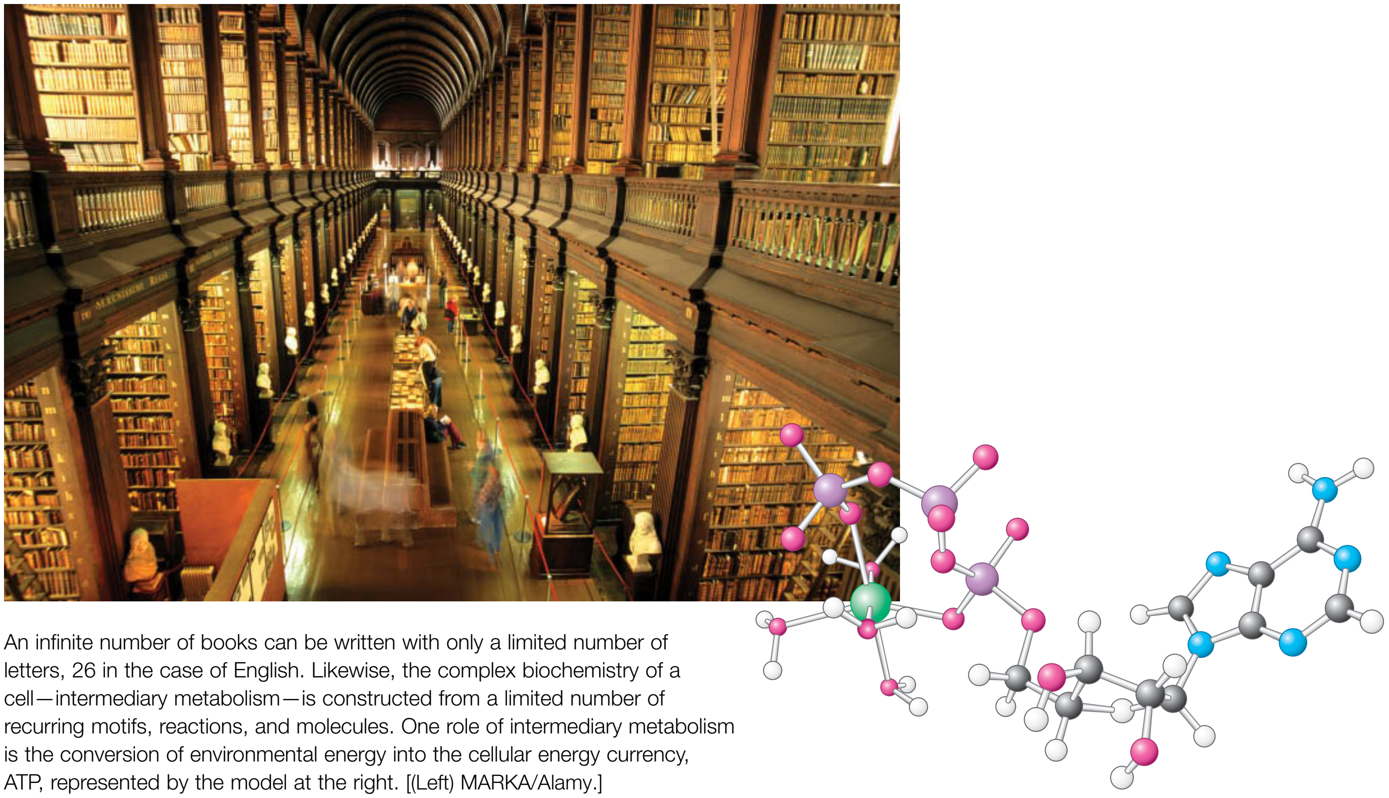
(b) What is the purpose of performing the experiment in the presence of excess nonradioactive ligand?

(c) What is the significance of the fact that specific binding attains a plateau?

23. *Counting receptors.* With the use of experiments such as those described in Problems 21 and 22, the number of receptors in the cell membrane can be calculated. Suppose that the specific activity of the ligand is 10^{12} cpm per millimole and that the maximal specific binding is 10^4 cpm per milligram of membrane protein. There are 10^{10} cells per milligram of membrane protein. Assume that one ligand binds per receptor. Calculate the number of receptor molecules present per cell.

CHAPTER
15

Metabolism: Basic Concepts and Design



An infinite number of books can be written with only a limited number of letters, 26 in the case of English. Likewise, the complex biochemistry of a cell—intermediary metabolism—is constructed from a limited number of recurring motifs, reactions, and molecules. One role of intermediary metabolism is the conversion of environmental energy into the cellular energy currency, ATP, represented by the model at the right. [(Left) MARKA/Alamy.]

OUTLINE

15.1 Metabolism Is Composed of Many Coupled, Interconnecting Reactions

15.2 ATP Is the Universal Currency of Free Energy in Biological Systems

15.3 The Oxidation of Carbon Fuels Is an Important Source of Cellular Energy

15.4 Metabolic Pathways Contain Many Recurring Motifs

The concepts of conformation and dynamics developed in Part I—especially those dealing with the specificity and catalytic power of enzymes, the regulation of their catalytic activity, and the transport of molecules and ions across membranes—enable us to now ask questions fundamental to biochemistry:

1. *How does a cell extract energy and reducing power from its environment?*

2. *How does a cell synthesize the building blocks of its macromolecules and then the macromolecules themselves?*

These processes are carried out by a highly integrated network of chemical reactions collectively known as *metabolism* or *intermediary metabolism*.

More than a thousand chemical reactions take place in even as simple an organism as *Escherichia coli*. The array of reactions may seem overwhelming at first glance. However, closer scrutiny reveals that metabolism has a *coherent design containing many common motifs*. These motifs include the use of an energy currency and the repeated appearance of a limited number of activated intermediates. In fact, a group of about 100 molecules play central roles in all forms of life. Furthermore, although the number of reactions in metabolism is large, the number of *kinds* of reactions is small and the mechanisms of these reactions are usually quite simple. Metabolic pathways are also regulated in common ways. The purpose of this chapter is to introduce some general principles of metabolism to provide a foundation for the more detailed studies to follow. These principles are:

1. Fuels are degraded and large molecules are constructed step by step in a series of linked reactions called *metabolic pathways*.
2. An energy currency common to all life forms, *adenosine triphosphate (ATP)*, links energy-releasing pathways with energy-requiring pathways.
3. The oxidation of carbon fuels powers the formation of ATP.
4. Although there are many metabolic pathways, a limited number of types of reactions and particular intermediates are common to many pathways.
5. Metabolic pathways are highly regulated.

15.1 Metabolism Is Composed of Many Coupled, Interconnecting Reactions

Living organisms require a continual input of free energy for three major purposes: (1) the performance of mechanical work in muscle contraction and cellular movements, (2) the active transport of molecules and ions, and (3) the synthesis of macromolecules and other biomolecules from simple precursors. The free energy used in these processes, which maintain an organism in a state that is far from equilibrium, is derived from the environment. Photosynthetic organisms, or *phototrophs,* obtain this energy by trapping sunlight, whereas *chemotrophs,* which include animals, obtain energy through the oxidation of foodstuffs generated by phototrophs.

Metabolism consists of energy-yielding and energy-requiring reactions

Metabolism is essentially a sequence of chemical reactions that begins with a particular molecule and results in the formation of some other molecule or molecules in a carefully defined fashion (Figure 15.1). There are many such defined pathways in the cell (Figure 15.2), and we will examine a few of them in some detail later. These pathways are interdependent, and their activity is coordinated by exquisitely sensitive means of communication in which allosteric enzymes are predominant (Section 10.1). We considered the principles of this communication in Chapter 14.

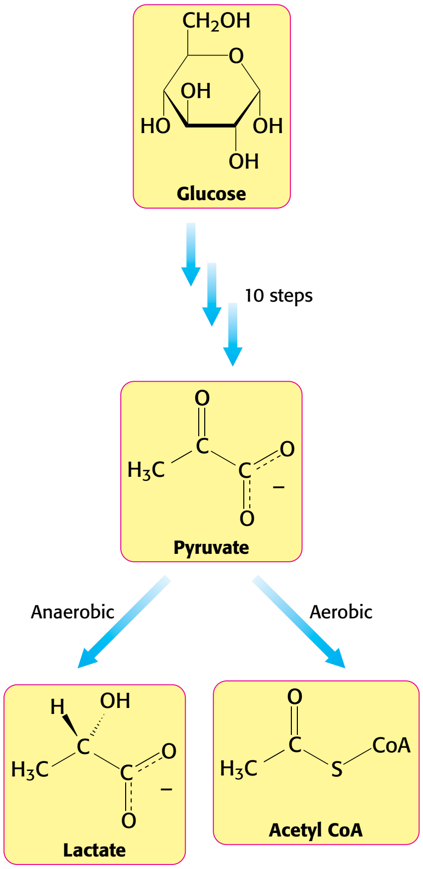


FIGURE 15.1 Glucose metabolism. Glucose is metabolized to pyruvate in 10 linked reactions. Under anaerobic conditions, pyruvate is metabolized to lactate and, under aerobic conditions, to acetyl CoA. The glucose-derived carbons of acetyl CoA are subsequently oxidized to CO_2.

We can divide metabolic pathways into two broad classes: (1) those that convert energy from fuels into biologically useful forms, and (2) those that require inputs of energy to proceed. Although this division is often imprecise, it is nonetheless a useful distinction in an examination of metabolism. Those reactions that transform fuels into cellular energy are called *catabolic reactions* or, more generally, *catabolism*.

$$\text{Fuel (carbohydrates, fats)} \xrightarrow{\text{Catabolism}} CO_2 + H_2O + \text{useful energy}$$

Those reactions that require energy—such as the synthesis of glucose, fats, or DNA—are called *anabolic reactions* or *anabolism*. The useful forms of energy that are produced in catabolism are employed in anabolism to generate complex structures from simple ones, or energy-rich states from energy-poor ones.

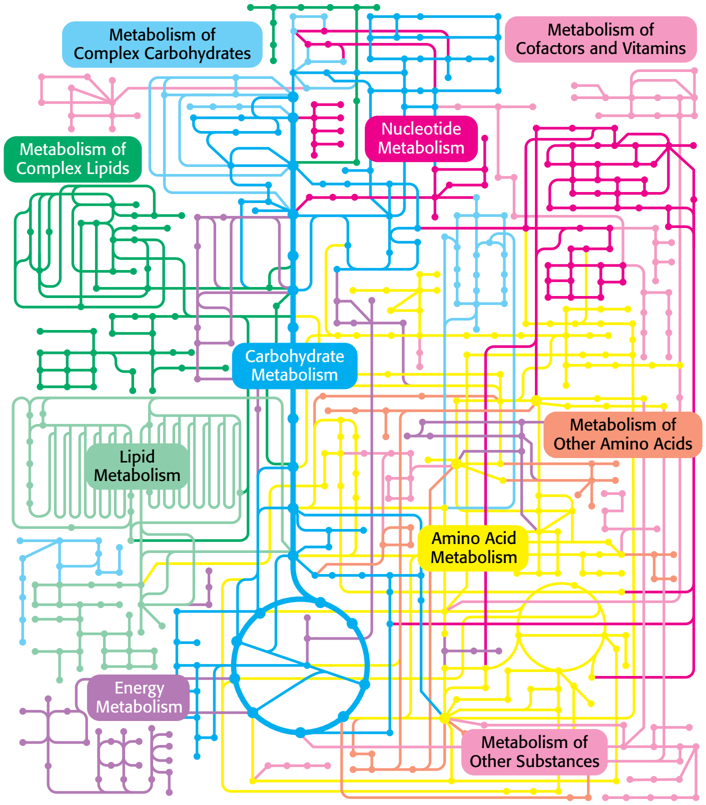

Metabolism of Complex Carbohydrates
Metabolism of Cofactors and Vitamins
Nucleotide Metabolism
Metabolism of Complex Lipids
Carbohydrate Metabolism
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metabolism
Energy Metabolism
Metabolism of Other Substances

FIGURE 15.2 Metabolic pathways. Each node represents a specific metabolite. [From the Kyoto Encyclopedia of Genes and Genomes (www.genome.ad.jp/kegg).]

$$\text{Useful energy} + \text{simple precursors} \xrightarrow{\text{Anabolism}} \text{complex molecules}$$

Some pathways can be either anabolic or catabolic, depending on the energy conditions in the cell. These pathways are referred to as *amphibolic pathways*.

An important general principle of metabolism is that *biosynthetic and degradative pathways are almost always distinct*. This separation is necessary for energetic reasons, as will be evident in subsequent chapters. It also facilitates the control of metabolism.

A thermodynamically unfavorable reaction can be driven by a favorable reaction

How are specific pathways constructed from individual reactions? A pathway must satisfy minimally two criteria: (1) the individual reactions must be *specific,* and (2) the entire set of reactions that constitute the pathway must be *thermodynamically favored*. A reaction that is specific will yield only one particular product or set of products from its reactants. As discussed in Chapter 8, enzymes provide this specificity. The thermodynamics of metabolism is most readily approached in relation to free energy, which was discussed in chapters 1 and 8. A reaction can occur spontaneously only if ΔG, the change in free energy, is negative. Recall that ΔG for the formation of products C and D from substrates A and B is given by

$$\Delta G = \Delta G^{\circ\prime} + RT \ln \frac{[\mathrm{C}][\mathrm{D}]}{[\mathrm{A}][\mathrm{B}]}$$

Thus, the ΔG of a reaction depends on the *nature* of the reactants and products (expressed by the $\Delta G^{\circ\prime}$ term, the standard free-energy change) and on their *concentrations* (expressed by the second term).

An important thermodynamic fact is that *the overall free-energy change for a chemically coupled series of reactions is equal to the sum of the free-energy changes of the individual steps*. Consider the following reactions:

$$\begin{array}{lll} \mathrm{A} \rightleftharpoons \mathrm{B} + \mathrm{C} & \Delta G^{\circ\prime} = +21\ \mathrm{kJ\ mol^{-1}}\ (+5\ \mathrm{kcal\ mol^{-1}}) \\ \mathrm{B} \rightleftharpoons \mathrm{D} & \Delta G^{\circ\prime} = -34\ \mathrm{kJ\ mol^{-1}}\ (-8\ \mathrm{kcal\ mol^{-1}}) \\ \hline \mathrm{A} \rightleftharpoons \mathrm{C} + \mathrm{D} & \Delta G^{\circ\prime} = -13\ \mathrm{kJ\ mol^{-1}}\ (-3\ \mathrm{kcal\ mol^{-1}}) \end{array}$$

Under standard conditions, A cannot be spontaneously converted into B and C, because $\Delta G^{\circ\prime}$ is positive. However, the conversion of B into D under standard conditions is thermodynamically feasible. Because free-energy changes are additive, the conversion of A into C and D has a $\Delta G^{\circ\prime}$ of -13 kJ $\mathrm{mol^{-1}}$ (-3 kcal $\mathrm{mol^{-1}}$), which means that it can occur spontaneously under standard conditions. Thus, *a thermodynamically unfavorable reaction can be driven by a thermodynamically favorable reaction to which it is coupled*. In this example, the reactions are coupled by the shared chemical intermediate B. Metabolic pathways are formed by the coupling of enzyme-catalyzed reactions such that the overall free energy of the pathway is negative.

15.2 ATP Is the Universal Currency of Free Energy in Biological Systems

Just as commerce is facilitated by the use of a common currency, the commerce of the cell—metabolism—is facilitated by the use of a common energy currency, *adenosine triphosphate* (ATP). Part of the free energy derived from the oxidation of foodstuffs and from light is transformed into this highly accessible molecule, which acts as the free-energy donor in most energy-requiring processes such as motion, active transport, and biosynthesis. Indeed, most of catabolism consists of reactions that extract energy from fuels such as carbohydrates and fats and convert it into ATP.

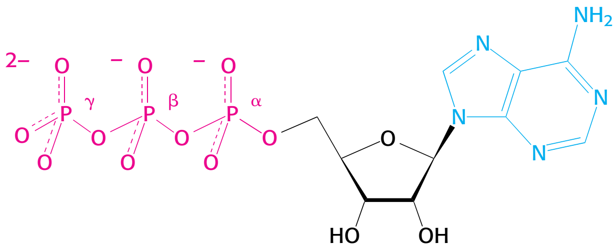
ATP hydrolysis is exergonic

ATP is a nucleotide consisting of adenine, a ribose, and a triphosphate unit (Figure 15.3). The active form of ATP is usually a complex of ATP with $\mathrm{Mg^{2+}}$ or $\mathrm{Mn^{2+}}$. In considering the role of ATP as an energy carrier, we can focus on its triphosphate moiety. *ATP is an energy-rich molecule because its triphosphate unit contains two phosphoanhydride bonds*. A large amount of free energy is liberated when ATP is hydrolyzed to adenosine diphosphate (ADP) and orthophosphate ($\mathrm{P_i}$) or when ATP is hydrolyzed to adenosine monophosphate (AMP) and pyrophosphate ($\mathrm{PP_i}$).

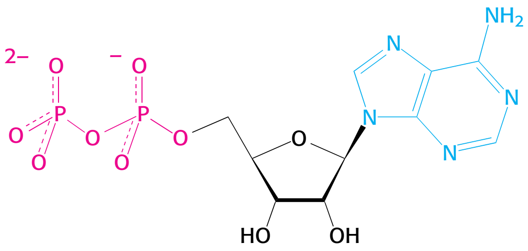
$$\mathrm{ATP} + \mathrm{H_2O} \rightleftharpoons \mathrm{ADP} + \mathrm{P_i}$$
$$\Delta G^{\circ\prime} = -30.5\ \mathrm{kJ\ mol^{-1}}\ (-7.3\ \mathrm{kcal\ mol^{-1}})$$

$$\mathrm{ATP} + \mathrm{H_2O} \rightleftharpoons \mathrm{AMP} + \mathrm{PP_i}$$
$$\Delta G^{\circ\prime} = -45.6\ \mathrm{kJ\ mol^{-1}}\ (-10.9\ \mathrm{kcal\ mol^{-1}})$$

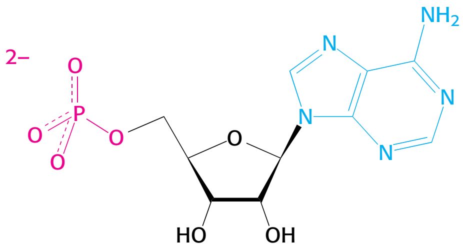
The precise ΔG for these reactions depends on the ionic strength of the medium and on the concentrations of $\mathrm{Mg^{2+}}$ and other metal ions (problems 23 and 34). Under typical cellular concentrations, the ΔG for these hydrolyses is approximately -50 kJ $\mathrm{mol^{-1}}$ (-12 kcal $\mathrm{mol^{-1}}$).



Adenosine triphosphate (ATP)



Adenosine diphosphate (ADP)



Adenosine monophosphate (AMP)

FIGURE 15.3 Structures of ATP, ADP, and AMP. These adenylates consist of adenine (blue), a ribose (black), and a tri-, di-, or monophosphate unit (red). The innermost phosphorus atom of ATP is designated P_α, the middle one P_β, and the outermost one P_γ.

The free energy liberated in the hydrolysis of ATP is harnessed to drive reactions that require an input of free energy, such as muscle contraction. In turn, ATP is formed from ADP and P_i when fuel molecules are oxidized in chemotrophs or when light is trapped by phototrophs. *This ATP–ADP cycle is the fundamental mode of energy exchange in biological systems.*

Some biosynthetic reactions are driven by the hydrolysis of other nucleoside triphosphates—namely, guanosine triphosphate (GTP), uridine triphosphate (UTP), and cytidine triphosphate (CTP). The diphosphate forms of these nucleotides are denoted by GDP, UDP, and CDP, and the monophosphate forms are denoted by GMP, UMP, and CMP. Enzymes catalyze the transfer of the terminal phosphoryl group from one nucleotide to another. The phosphorylation of nucleoside monophosphates is catalyzed by a family of *nucleoside monophosphate kinases,* as discussed in Section 9.4. The phosphorylation of nucleoside diphosphates is catalyzed by *nucleoside diphosphate kinase,* an enzyme with broad specificity.

$$\underset{\text{Nucleoside monophosphate}}{\text{NMP}} + \text{ATP} \xrightleftharpoons{\text{Nucleoside monophosphate kinase}} \text{NDP} + \text{ADP}$$

$$\underset{\text{Nucleoside diphosphate}}{\text{NDP}} + \text{ATP} \xrightleftharpoons{\text{Nucleoside diphosphate kinase}} \text{NTP} + \text{ADP}$$

It is intriguing to note that although all of the nucleotide triphosphates are energetically equivalent, ATP is nonetheless the primary cellular energy carrier. In addition, two important electron carriers, NAD^+ and FAD, as well the acyl group carrier, coenzyme A, are derivatives of ATP. *The role of ATP in energy metabolism is paramount.*

ATP hydrolysis drives metabolism by shifting the equilibrium of coupled reactions

An otherwise unfavorable reaction can be made possible by coupling to ATP hydrolysis. Consider a reaction that is thermodynamically unfavorable without an input of free energy, a situation common to most biosynthetic reactions. Suppose that the standard free energy of the conversion of compound A into compound B is $+16.7\ \text{kJ mol}^{-1}$ ($+4.0\ \text{kcal mol}^{-1}$):

$$\text{A} \rightleftharpoons \text{B} \qquad \Delta G^{\circ\prime} = +16.7\ \text{kJ mol}^{-1}\ (+4\ \text{kcal mol}^{-1})$$

The equilibrium constant K'_{eq} of this reaction at 25°C is related to $\Delta G^{\circ\prime}$ (in units of kilojoules per mole) by

$$K'_{\text{eq}} = [\text{B}]_{\text{eq}}/[\text{A}]_{\text{eq}} = e^{-\Delta G^{\circ\prime}/2.47} = 1.15 \times 10^{-3}$$

Thus, net conversion of A into B cannot take place when the molar ratio of B to A is equal to or greater than 1.15×10^{-3}. However, A can be converted into B under these conditions if the reaction is coupled to the hydrolysis of ATP. Under standard conditions, the $\Delta G^{\circ\prime}$ of hydrolysis is approximately $-30.5\ \text{kJ mol}^{-1}$ ($-7.3\ \text{kcal mol}^{-1}$). The new overall reaction is

$$\text{A} + \text{ATP} + \text{H}_2\text{O} \rightleftharpoons \text{B} + \text{ADP} + \text{P}_\text{i}$$
$$\Delta G^{\circ\prime} = -13.8\ \text{kJ mol}^{-1}(-3.3\ \text{kcal mol}^{-1})$$

Its free-energy change of $-13.8\ \text{kJ mol}^{-1}$ ($-3.3\ \text{kcal mol}^{-1}$) is the sum of the value of $\Delta G^{\circ\prime}$ for the conversion of A into B [$+16.7\ \text{kJ mol}^{-1}$ ($+4.0\ \text{kcal mol}^{-1}$)] and the value of $\Delta G^{\circ\prime}$ for the hydrolysis of ATP [$-30.5\ \text{kJ mol}^{-1}$ ($-7.3\ \text{kcal mol}^{-1}$)]. At pH 7, the equilibrium constant of this coupled reaction is

$$K'_{\text{eq}} = \frac{[\text{B}]_{\text{eq}}}{[\text{A}]_{\text{eq}}} \times \frac{[\text{ADP}]_{\text{eq}}[\text{P}_\text{i}]_{\text{eq}}}{[\text{ATP}]_{\text{eq}}} = e^{13.8/2.47} = 2.67 \times 10^{2}$$

At equilibrium, the ratio of [B] to [A] is given by

$$\frac{[\text{B}]_{\text{eq}}}{[\text{A}]_{\text{eq}}} = K'_{\text{eq}} \frac{[\text{ATP}]_{\text{eq}}}{[\text{ADP}]_{\text{eq}}[\text{P}_\text{i}]_{\text{eq}}}$$

which means that the hydrolysis of ATP enables A to be converted into B until the [B]/[A] ratio reaches a value of 2.67×10^{2}. This equilibrium ratio is strikingly different from the value of 1.15×10^{-3} for the reaction $\text{A} \rightarrow \text{B}$ in the absence of ATP hydrolysis. In other words, coupling the hydrolysis of ATP with the conversion of A into B under standard conditions has changed the equilibrium ratio of B to A by a factor of about 10^{5}. If we were to use the ΔG of hydrolysis of ATP under cellular conditions [$-50.2\ \text{kJ mol}^{-1}$ ($-12\ \text{kcal mol}^{-1}$)] in our calculations instead of $\Delta G^{\circ\prime}$, the change in the equilibrium ratio would be even more dramatic, on the order of 10^{8}.

We see here the thermodynamic essence of ATP's action as an *energy-coupling agent*. Cells maintain ATP levels by using oxidizable substrates or light as sources of free energy for synthesizing the molecule. In the cell, the hydrolysis of an ATP molecule in a coupled reaction then changes the equilibrium ratio of products to reactants by a very large factor, of the order of 10^{8}. More generally, the hydrolysis of n ATP molecules changes the equilibrium ratio of a coupled reaction (or sequence of reactions) by a factor of 10^{8n}. For example, the hydrolysis of three ATP molecules in a coupled reaction changes the equilibrium ratio by a factor of 10^{24}. Thus, *a thermodynamically unfavorable reaction sequence can be converted into a favorable one by coupling it to the hydrolysis of a sufficient number of ATP molecules in a new reaction*. It should also be emphasized that A and B in the preceding coupled reaction may be interpreted very generally, not only as different chemical species. For example, A and B may represent activated and unactivated conformations of a protein that is activated by phosphorylation with ATP.

Through such changes in protein conformation, molecular motors such as myosin, kinesin, and dynein convert the chemical energy of ATP into mechanical energy (Chapter 34). Indeed, this conversion is the basis of muscle contraction.

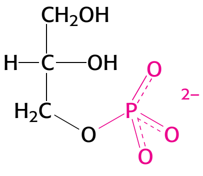
Alternatively, A and B may refer to the concentrations of an ion or molecule on the outside and inside of a cell, as in the active transport of a nutrient. The active transport of Na^+ and K^+ across membranes is driven by the phosphorylation of the sodium–potassium pump by ATP and its subsequent dephosphorylation (Section 13.2).

The high phosphoryl potential of ATP results from structural differences between ATP and its hydrolysis products

What makes ATP an efficient phosphoryl-group donor? Let us compare the standard free energy of hydrolysis of ATP with that of a phosphate ester, such as glycerol 3-phosphate:

$$\text{ATP} + H_2O \rightleftharpoons \text{ADP} + P_i$$
$$\Delta G^{\circ\prime} = -30.5\ \text{kJ mol}^{-1}\ (-7.3\ \text{kcal mol}^{-1})$$

$$\text{Glycerol 3-phosphate} + H_2O \rightleftharpoons \text{glycerol} + P_i$$
$$\Delta G^{\circ\prime} = -9.2\ \text{kJ mol}^{-1}\ (-2.2\ \text{kcal mol}^{-1})$$



Glycerol 3-phosphate

The magnitude of $\Delta G^{\circ\prime}$ for the hydrolysis of glycerol 3-phosphate is much smaller than that of ATP, which means that ATP has a stronger tendency to transfer its terminal phosphoryl group to water than does glycerol 3-phosphate. In other words, ATP has a higher *phosphoryl-transfer potential (phosphoryl-group-transfer potential)* than does glycerol 3-phosphate.

The high phosphoryl-transfer potential of ATP can be explained by features of the ATP structure. Because $\Delta G^{\circ\prime}$ depends on the *difference* in free energies of the products and reactants, we need to examine the structures of both ATP and its hydrolysis products, ADP and P_i, to answer this question. Four factors are important: *resonance stabilization, electrostatic repulsion, increase in entropy,* and *stabilization due to hydration.*

1. *Resonance Stabilization.* Orthophosphate (P_i), one of the products of ATP hydrolysis, has greater resonance stabilization than do any of the phosphoryl groups of ATP. Orthophosphate has a number of resonance forms of similar energy (Figure 15.4), whereas the γ phosphoryl group of ATP has a smaller number. Forms like that shown in Figure 15.5 are unfavorable because a positively charged oxygen atom is adjacent to a positively charged phosphorus atom, an electrostatically unfavorable juxtaposition.

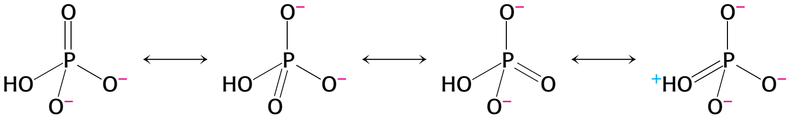


FIGURE 15.4 Resonance structures of orthophosphate.

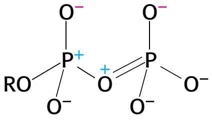


FIGURE 15.5 Improbable resonance structure. The structure contributes little to the terminal part of ATP, because two positive charges are placed adjacent to each other.

2. *Electrostatic Repulsion.* At pH 7, the triphosphate unit of ATP carries about four negative charges. These charges repel one another because they are in close proximity. The repulsion between them is reduced when ATP is hydrolyzed.

3. *Increase in Entropy*. The entropy of the products of ATP hydrolysis is greater, in that there are now two molecules instead of a single ATP molecule. We disregard the molecule of water used to hydrolyze the ATP; given the high concentration (55.5 M), there is effectively no change in the concentration of water during the reaction.

4. *Stabilization Due to Hydration.* Water binds to ADP and P_i, stabilizing these molecules, and rendering the reverse reaction, the synthesis of ATP, more unfavorable.

ATP is often called a high-energy phosphate compound, and its phosphoanhydride bonds are referred to as high-energy bonds. Indeed, a "squiggle" (~P) is often used to indicate such a bond. Nonetheless, there is nothing special about the bonds themselves. *They are high-energy bonds in the sense that much free energy is released when they are hydrolyzed*, for the reasons listed above.

Phosphoryl-transfer potential is an important form of cellular energy transformation

The standard free energies of hydrolysis provide a convenient means of comparing the phosphoryl-transfer potential of phosphorylated compounds. Such comparisons reveal that ATP is not the only compound with a high phosphoryl-transfer potential. In fact, some compounds in biological systems have a higher phosphoryl-transfer potential than that of ATP. These compounds include phosphoenolpyruvate (PEP), 1,3-bisphosphoglycerate (1,3-BPG), and creatine phosphate (Figure 15.6). Thus, PEP can transfer its phosphoryl group to ADP to form ATP. Indeed, this transfer is one of the ways in which ATP is generated in the breakdown of sugars (Chapter 16). It is significant that ATP has a phosphoryl-transfer potential that is intermediate among the biologically important phosphorylated molecules (Table 15.1). *This intermediate position enables ATP to function efficiently as a carrier of phosphoryl groups.*

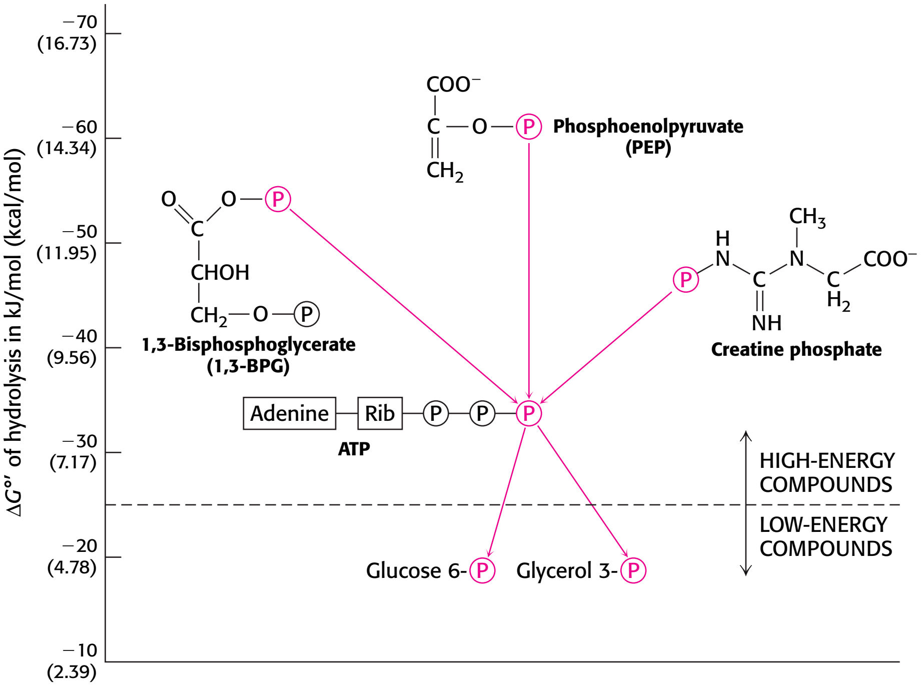


FIGURE 15.6 Compounds with high phosphoryl-transfer potential. The role of ATP as the cellular energy currency is illustrated by its relation to other phosphorylated compounds. ATP has a phosphoryl-transfer potential that is intermediate among the biologically important phosphorylated molecules. High-phosphoryl-transfer-potential compounds (1,3-BPG, PEP, and creatine phosphate) derived from the metabolism of fuel molecules are used to power ATP synthesis. In turn, ATP donates a phosphoryl group to other biomolecules to facilitate their metabolism. [Data from D. L. Nelson and M. M. Cox, Lehninger Principles of Biochemistry, 5th ed. (W. H. Freeman and Company, 2009), Fig. 13-19.]

TABLE 15.1 Standard free energies of hydrolysis of some phosphorylated compounds

Compound	kJ mol^{-1}	kcal mol^{-1}
Phosphoenolpyruvate	−61.9	−14.8
1,3-Bisphosphoglycerate	−49.4	−11.8
Creatine phosphate	−43.1	−10.3
ATP (to ADP)	−30.5	− 7.3
Glucose 1-phosphate	−20.9	− 5.0
Pyrophosphate	−19.3	− 4.6
Glucose 6-phosphate	−13.8	− 3.3
Glycerol 3-phosphate	− 9.2	− 2.2

The amount of ATP in muscle suffices to sustain contractile activity for less than a second. Creatine phosphate in vertebrate muscle serves as a reservoir of high-potential phosphoryl groups that can be readily transferred to ADP. Indeed, we use creatine phosphate to regenerate ATP from ADP every time that we exercise strenuously. This reaction is catalyzed by *creatine kinase*.

$$\text{Creatine phosphate} + \text{ADP} \xrightleftharpoons{\text{Creatine kinase}} \text{ATP} + \text{creatine}$$

At pH 7, the standard free energy of hydrolysis of creatine phosphate is −43.1 kJ mol^{-1} (−10.3 kcal mol^{-1}), compared with −30.5 kJ mol^{-1} (−7.3 kcal mol^{-1}) for ATP. Hence, the standard free-energy change in forming ATP from creatine phosphate is −12.6 kJ mol^{-1} (−3.0 kcal mol^{-1}), which corresponds to an equilibrium constant of 162.

$$K_{\text{eq}} = \frac{[\text{ATP}][\text{creatine}]}{[\text{ADP}][\text{creatine phosphate}]} = e^{-\Delta G^{\circ\prime}/2.47} = e^{12.6/2.47} = 162$$

In resting muscle, typical concentrations of these metabolites are [ATP] = 4 mM, [ADP] = 0.013 mM, [creatine phosphate] = 25 mM, and [creatine] = 13 mM. Because of its abundance and high phosphoryl-transfer potential relative to that of ATP, creatine phosphate is a highly effective phosphoryl buffer. Indeed, creatine phosphate is the major source of phosphoryl groups for ATP regeneration for a runner during the first 4 seconds of a 100-meter sprint. The fact that creatine phosphate can replenish ATP pools is the basis of the use of creatine as a dietary supplement by athletes in sports requiring short bursts of intense activity. After the creatine phosphate pool is depleted, ATP must be generated through metabolism (Figure 15.7).

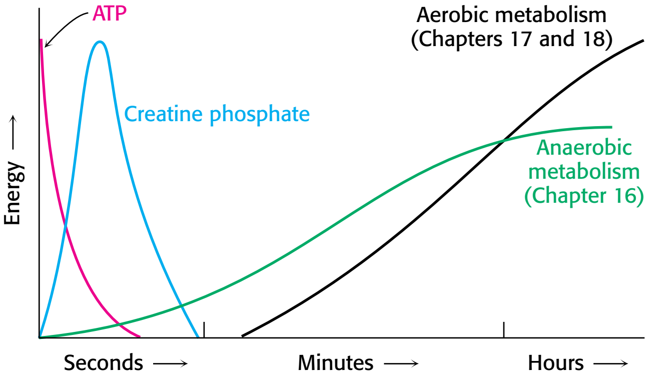


FIGURE 15.7 Sources of ATP during exercise. In the initial seconds, exercise is powered by existing high-phosphoryl-transfer compounds (ATP and creatine phosphate). Subsequently, the ATP must be regenerated by metabolic pathways.

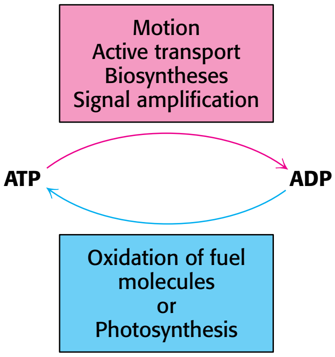
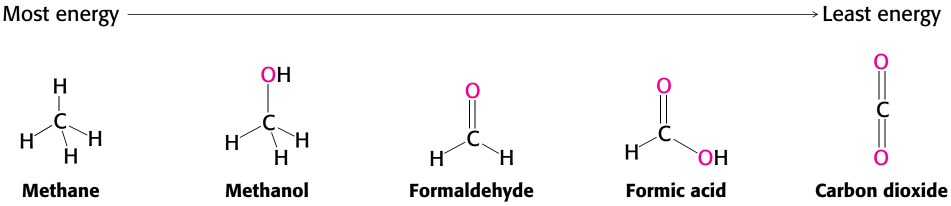


FIGURE 15.8 ATP–ADP cycle. This cycle is the fundamental mode of energy exchange in biological systems.

15.3 The Oxidation of Carbon Fuels Is an Important Source of Cellular Energy

ATP serves as the principal *immediate donor of free energy* in biological systems rather than as a long-term storage form of free energy. In a typical cell, an ATP molecule is consumed within a minute of its formation. Although the total quantity of ATP in the body is limited to approximately 100 g, *the turnover of this small quantity of ATP is very high*. For example, a resting human being consumes about 40 kg (88 pounds) of ATP in 24 hours. During strenuous exertion, the rate of utilization of ATP may be as high as 0.5 kg/minute. For a 2-hour run, 60 kg (132 pounds) of ATP is utilized. Clearly, having mechanisms for regenerating ATP is vital. Motion, active transport, signal amplification, and biosynthesis can take place only if ATP is continually regenerated from ADP (Figure 15.8). *The generation of ATP is one of the primary roles of catabolism.* The carbon in fuel molecules—such as glucose and fats—is oxidized to CO_2. The resulting electrons are captured and used to regenerate ATP from ADP and P_i.

In aerobic organisms, the ultimate electron acceptor in the oxidation of carbon is O_2 and the oxidation product is CO_2. Consequently, the more reduced a carbon is to begin with, the more free energy is released by its oxidation. Figure 15.9 shows the $\Delta G^{\circ\prime}$ of oxidation for one-carbon compounds.



	Methane	Methanol	Formaldehyde	Formic acid	Carbon dioxide
$\Delta G^{\circ\prime}_{\text{oxidation}}$ (kJ mol^{-1})	−820	−703	−523	−285	0
$\Delta G^{\circ\prime}_{\text{oxidation}}$ (kcal mol^{-1})	−196	−168	−125	−68	0

FIGURE 15.9 Free energy of oxidation of single-carbon compounds.

Fuel molecules are more complex (Figure 15.10) than the single-carbon compounds depicted in Figure 15.9. Nevertheless, oxidation of these fuels takes place one carbon at a time. The carbon-oxidation energy is used in some cases to create a compound with high phosphoryl-transfer potential and in other cases to create an ion gradient. In either case, the end point is the formation of ATP.

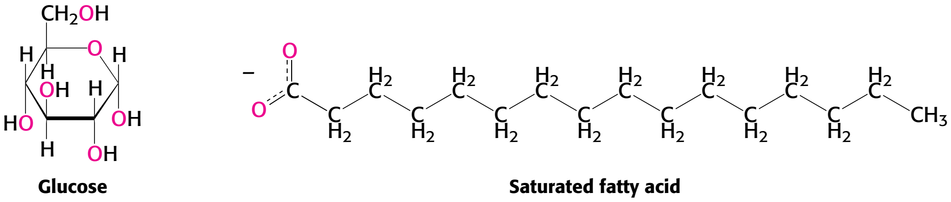
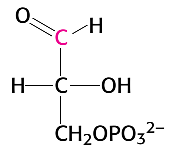
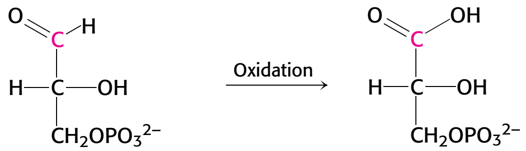


FIGURE 15.10 Prominent fuels. Fats are a more efficient fuel source than carbohydrates such as glucose because the carbon in fats is more reduced.

Compounds with high phosphoryl-transfer potential can couple carbon oxidation to ATP synthesis

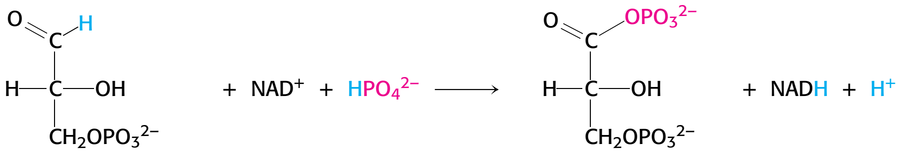
How is the energy released from the oxidation of a carbon compound converted into ATP? As an example, consider glyceraldehyde 3-phosphate (shown in the margin), which is a metabolite of glucose formed in the oxidation of that sugar. The C-1 carbon (shown in red) is at the aldehyde-oxidation level and is not in its most oxidized state. Oxidation of the aldehyde to an acid will release energy.


Glyceraldehyde 3-phosphate (GAP)



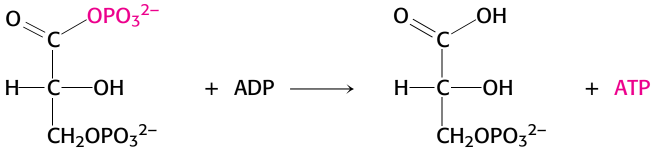
Glyceraldehyde 3-phosphate **3-Phosphoglyceric acid**

However, the oxidation does not take place directly. Instead, the carbon oxidation generates an acyl phosphate, 1,3-bisphosphoglycerate. The electrons released are captured by NAD$^+$, which we will consider shortly.



Glyceraldehyde 3-phosphate (GAP) **1,3-Bisphosphoglycerate (1,3-BPG)**

For reasons similar to those discussed for ATP, 1,3-bisphosphoglycerate has a high phosphoryl-transfer potential that is, in fact, greater than that of ATP. Thus, the hydrolysis of 1,3-BPG can be coupled to the synthesis of ATP.



1,3-Bisphosphoglycerate **3-Phosphoglyceric acid**

The energy of oxidation is initially trapped as a high-phosphoryl-transfer-potential compound and then used to form ATP. The oxidation energy of a carbon atom is transformed into phosphoryl-transfer potential, first as 1,3-bisphosphoglycerate and ultimately as ATP. We will consider these reactions in mechanistic detail in Chapter 16.

Ion gradients across membranes provide an important form of cellular energy that can be coupled to ATP synthesis

As described in Chapter 13, electrochemical potential is an effective means of storing free energy. Indeed, the electrochemical potential of *ion gradients across membranes,* produced by the oxidation of fuel molecules or by photosynthesis, ultimately powers the synthesis of most of the ATP in cells. In general, ion gradients are versatile means of coupling thermodynamically unfavorable reactions to favorable ones. Indeed, in animals, *proton gradients* generated by the oxidation of carbon fuels account for more than 90% of ATP generation (Figure 15.11). This process is called *oxidative phosphorylation* (Chapter 18). ATP hydrolysis can then be used to form ion gradients of different types and functions. The electrochemical potential of a

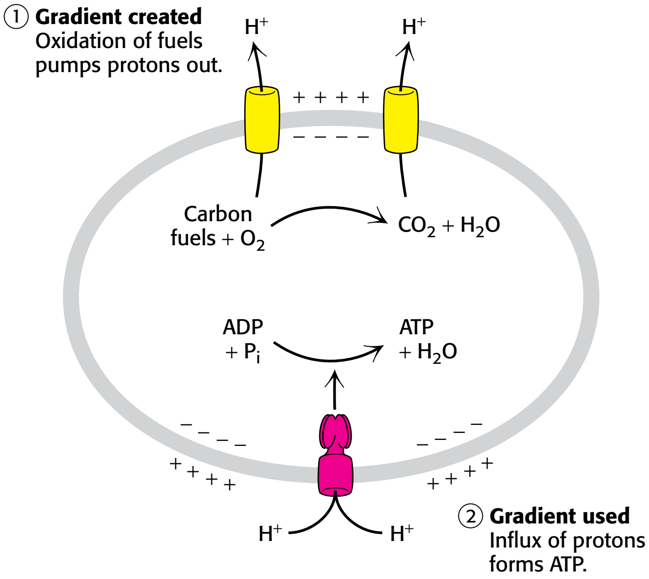


FIGURE 15.11 Proton gradients. The oxidation of fuels can power the formation of proton gradients by the action of specific proton pumps (yellow cylinders). These proton gradients can in turn drive the synthesis of ATP when the protons flow through an ATP-synthesizing enzyme (red complex).

Na^+ gradient, for example, can be tapped to pump Ca^{2+} out of cells or to transport nutrients such as sugars and amino acids into cells.

Phosphates play a prominent role in biochemical processes

We have seen in Section 10.3, Chapter 14, and in this chapter the prominence of phosphoryl group transfer from ATP to acceptor molecules. How is it that phosphate came to play such a prominent role in biology? Phosphate and its esters have several characteristics that render it useful for biochemical systems. First, phosphate esters have the important property of being *thermodynamically unstable* while being *kinetically stable*. Phosphate esters are thus molecules whose energy release can be manipulated by enzymes. The stability of phosphate esters is due to the negative charges that make them resistant to hydrolysis in the absence of enzymes. This accounts for the presence of phosphate in the backbone of DNA. Furthermore, because phosphate esters are so kinetically stable, they make ideal regulatory molecules, added to proteins by kinases and removed only by phosphatases. Phosphates are also frequently added to metabolites that might otherwise diffuse through the cell membrane. Furthermore, even when transporters exist for unphosphorylated forms of a metabolite, the addition of a phosphate changes the geometry and polarity of the molecules so that they no longer fit in the binding sites of the transporters.

No other ions have the chemical characteristics of phosphate. Citrate is not sufficiently charged to prevent hydrolysis. Arsenate forms esters that are unstable and susceptible to spontaneous hydrolysis. Indeed, arsenate is poisonous to cells because it can replace phosphate in reactions required for ATP synthesis, generating unstable compounds and preventing ATP synthesis. Silicate is more abundant than phosphate, but silicate salts are virtually insoluble, and in fact, are used for biomineralization. Only phosphate has the chemical properties to meet the needs of living systems.

Energy from foodstuffs is extracted in three stages

Let us take an overall view of the processes of energy conversion in higher organisms before considering them in detail in subsequent chapters. Hans Krebs described three stages in the generation of energy from the oxidation of foodstuffs (Figure 15.12).

In the first stage, large molecules in food are broken down into smaller units in the process of *digestion*. Proteins are hydrolyzed to their 20 different amino acids, polysaccharides are hydrolyzed to simple sugars such as glucose, and lipids are hydrolyzed to glycerol and fatty acids. The degradation products are then absorbed by the cells of the intestine and distributed throughout the body. This stage is strictly a preparation stage; no useful energy is captured in this phase.

In the second stage, these numerous small molecules are degraded to a few simple units that play a central role in metabolism. In fact, most of them—sugars, fatty acids, glycerol, and several amino acids—are converted into the acetyl unit of acetyl CoA. Some ATP is generated in this stage, but the amount is small compared with that obtained in the third stage.

In the third stage, ATP is produced from the complete oxidation of the acetyl unit of acetyl CoA. The third stage consists of the citric acid cycle and oxidative phosphorylation, which are *the final common pathways in the oxidation of fuel molecules.* Acetyl CoA brings acetyl units into the citric acid cycle [also called the tricarboxylic acid (TCA) cycle or Krebs cycle], where they are

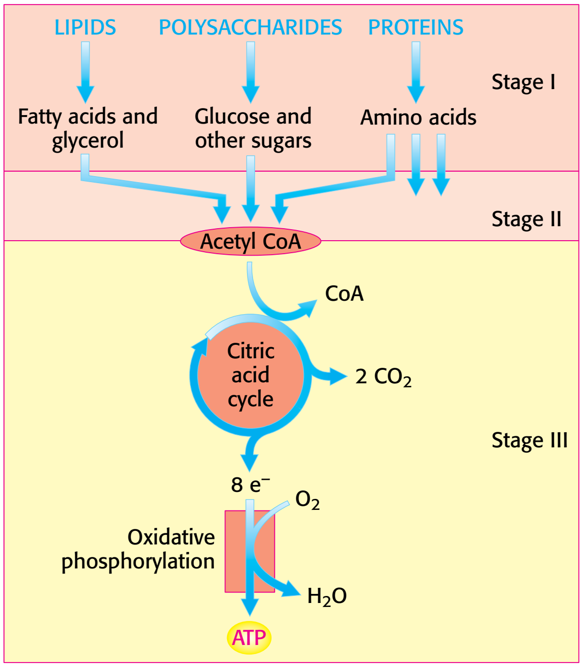


FIGURE 15.12 Stages of catabolism. The extraction of energy from fuels can be divided into three stages.

completely oxidized to CO_2. Four pairs of electrons are transferred (three to NAD^+ and one to FAD) for each acetyl group that is oxidized. Then, a proton gradient is generated as electrons flow from the reduced forms of these carriers to O_2, and this gradient is used to synthesize ATP (Chapters 17 and 18).

15.4 Metabolic Pathways Contain Many Recurring Motifs

At first glance, metabolism appears intimidating because of the sheer number of reactants and reactions. Nevertheless, there are unifying themes that make the comprehension of this complexity more manageable. These unifying themes include common metabolites, reactions, and regulatory schemes that stem from a common evolutionary heritage.

Activated carriers exemplify the modular design and economy of metabolism

We have seen that phosphoryl transfer can be used to drive otherwise endergonic reactions, alter the energy of conformation of a protein, or serve as a signal to alter the activity of a protein. The phosphoryl-group donor in all of these reactions is ATP. In other words, *ATP is an activated carrier of phosphoryl groups because phosphoryl transfer from ATP is an exergonic process*. The use of activated carriers is a recurring motif in biochemistry, and we will consider several such carriers here. Many such activated carriers function as coenzymes (Section 8.1):

1. *Activated Carriers of Electrons for Fuel Oxidation*. In aerobic organisms, the ultimate electron acceptor in the oxidation of fuel molecules is O_2. However, electrons are not transferred directly to O_2. Instead, fuel molecules transfer electrons to special carriers, which are either *pyridine nucleotides* or *flavins*. The reduced forms of these carriers then transfer their high-potential electrons to O_2.

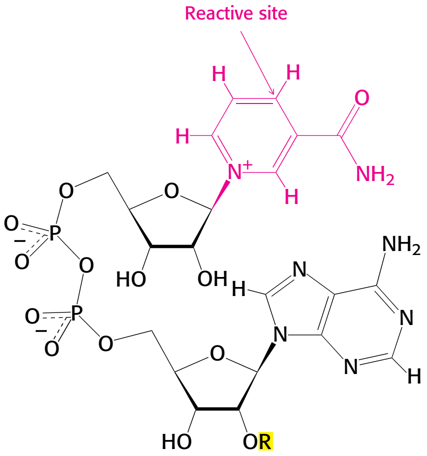
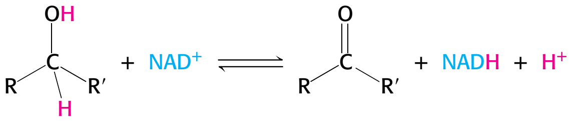


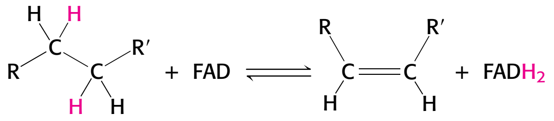
FIGURE 15.13 Structures of the oxidized forms of nicotinamide-derived electron carriers. Nicotinamide adenine dinucleotide (NAD^+) and nicotinamide adenine dinucleotide phosphate ($NADP^+$) are prominent carriers of high-energy electrons. In NAD^+, R = H; in $NADP^+$, $R = PO_3^{2-}$.

Nicotinamide adenine dinucleotide is a major electron carrier in the oxidation of fuel molecules (Figure 15.13). The reactive part of NAD^+ is its nicotinamide ring, a pyridine derivative synthesized from the vitamin niacin. *In the oxidation of a substrate, the nicotinamide ring of NAD^+ accepts a hydrogen ion and two electrons, which are equivalent to a hydride ion* ($H:^-$). The reduced form of this carrier is called *NADH*. In the oxidized form, the nitrogen atom carries a positive charge, as indicated by NAD^+. NAD^+ is the electron acceptor in many reactions of the type



In this dehydrogenation, one hydrogen atom of the substrate is directly transferred to NAD^+, whereas the other appears in the solvent as a proton. Both electrons lost by the substrate are transferred to the nicotinamide ring.

The other major electron carrier in the oxidation of fuel molecules is the coenzyme *flavin adenine dinucleotide* (Figure 15.14). The abbreviations for the oxidized and reduced forms of this carrier are FAD and $FADH_2$, respectively. FAD is the electron acceptor in reactions of the type



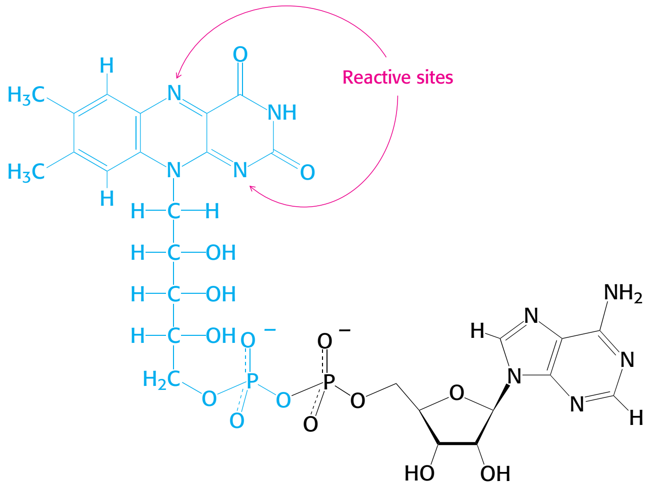


FIGURE 15.14 Structure of the oxidized form of flavin adenine dinucleotide (FAD). This electron carrier consists of a flavin mononucleotide (FMN) unit (shown in blue) and an AMP unit (shown in black).

The reactive part of FAD is its isoalloxazine ring, a derivative of the vitamin riboflavin (Figure 15.15). FAD, like NAD^+, can accept two electrons. In doing so, FAD, unlike NAD^+, takes up two protons. These carriers of high-potential electrons as well as flavin mononucleotide (FMN), an electron carrier similar to FAD but lacking the adenine nucleotide, will be considered further in Chapter 18.

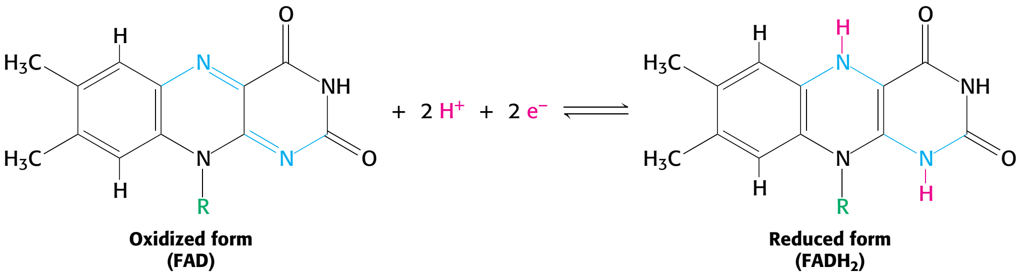
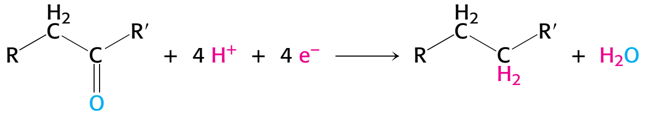


FIGURE 15.15 Structures of the reactive components of FAD and $FADH_2$. The electrons and protons are carried by the isoalloxazine ring component of FAD and $FADH_2$.

2. *An Activated Carrier of Electrons for Reductive Biosynthesis.* High-potential electrons are required in most biosyntheses because the precursors are more oxidized than the products. Hence, reducing power is needed in addition to ATP. For example, in the biosynthesis of fatty acids, a keto group is reduced to a methylene group in several steps. This sequence of reactions requires an input of four electrons.



The electron donor in most reductive biosyntheses is NADPH, the reduced form of nicotinamide adenine dinucleotide phosphate ($NADP^+$; Figure 15.13). NADPH differs from NADH in that the 2′-hydroxyl group of its adenosine moiety is esterified with phosphate. NADPH carries electrons in the same way as NADH. However, *NADPH is used almost exclusively for reductive biosyntheses, whereas NADH is used primarily for the generation of ATP*. The extra phosphoryl group on NADPH is a tag that enables enzymes to distinguish between high-potential electrons to be used in anabolism and those to be used in catabolism.

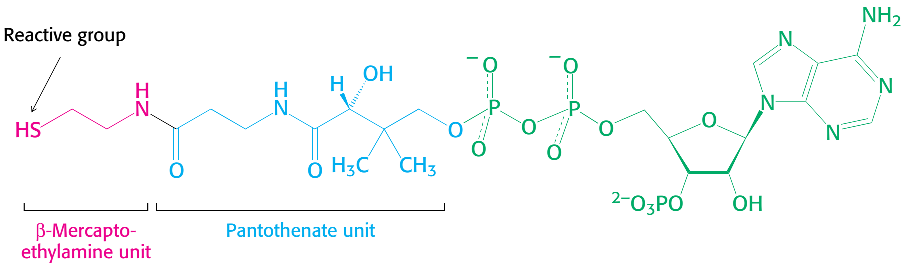
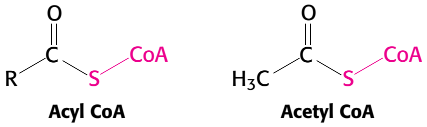


FIGURE 15.16 Structure of coenzyme A (CoA-SH).

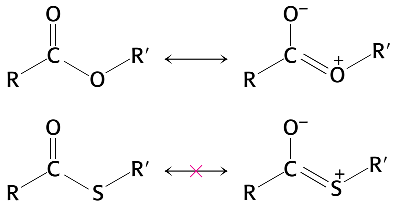
3. *An Activated Carrier of Two-Carbon Fragments*. Coenzyme A, another central molecule in metabolism, is a carrier of acyl groups derived from the vitamin pantothenate (Figure 15.16). Acyl groups are important constituents both in catabolism, as in the oxidation of fatty acids, and in anabolism, as in the synthesis of membrane lipids. The terminal sulfhydryl group in CoA is the reactive site. Acyl groups are linked to CoA by thioester bonds. The resulting derivative is called an *acyl CoA*. An acyl group often linked to CoA is the acetyl unit; this derivative is called *acetyl CoA*. The $\Delta G^{\circ\prime}$ for the hydrolysis of acetyl CoA has a large negative value:

$$\text{Acetyl CoA} + H_2O \rightleftharpoons \text{acetate} + \text{CoA} + H^+$$

$$\Delta G^{\circ\prime} = -31.4 \text{ kJ mol}^{-1}(-7.5 \text{ kcal mol}^{-1})$$



A thioester is thermodynamically more unstable than an oxygen ester because the electrons of the C═O bond cannot form resonance structures with the C—S bond that are as stable as those that they can form with the C—O bond. Consequently, *acetyl CoA has a high acetyl-group-transfer potential because transfer of the acetyl group is exergonic*. Acetyl CoA carries an activated acetyl group, just as ATP carries an activated phosphoryl group.



Oxygen esters are stabilized by resonance structures not available to thioesters.

The use of activated carriers illustrates two key aspects of metabolism. First, NADH, NADPH, and $FADH_2$ react slowly with O_2 in the absence of a catalyst. Likewise, ATP and acetyl CoA are hydrolyzed slowly (over many hours or even days) in the absence of a catalyst. These molecules are kinetically quite stable in the face of a large thermodynamic driving force for reaction with O_2 (in regard to the electron carriers) and H_2O (for ATP and acetyl CoA). *The kinetic stability of these molecules in the absence of specific catalysts is essential for their biological function because it enables enzymes to control the flow of free energy and reducing power*.

Second, *most interchanges of activated groups in metabolism are accomplished by a rather small set of carriers* (Table 15.2). The existence of a recurring set of

TABLE 15.2 Some activated carriers in metabolism

Carrier molecule in activated form	Group carried	Vitamin precursor
ATP	Phosphoryl	
NADH and NADPH	Electrons	Nicotinate (niacin) (vitamin B_3)
$FADH_2$	Electrons	Riboflavin (vitamin B_2)
$FMNH_2$	Electrons	Riboflavin (vitamin B_2)
Coenzyme A	Acyl	Pantothenate (vitamin B_5)
Lipoamide	Acyl	
Thiamine pyrophosphate	Aldehyde	Thiamine (vitamin B_1)
Biotin	CO_2	Biotin (vitamin B_7)
Tetrahydrofolate	One-carbon units	Folate (vitamin B_9)
S-Adenosylmethionine	Methyl	
Uridine diphosphate glucose	Glucose	
Cytidine diphosphate diacylglycerol	Phosphatidate	
Nucleoside triphosphates	Nucleotides	

Note: Many of the activated carriers are coenzymes that are derived from water-soluble vitamins.

TABLE 15.3 The B vitamins

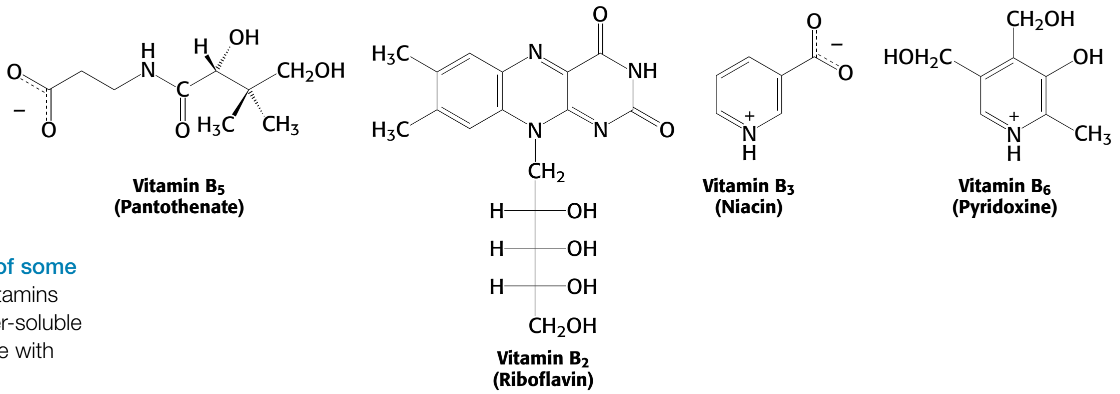
Vitamin	Coenzyme	Typical reaction type	Consequences of deficiency
Thiamine (B_1)	Thiamine pyrophosphate	Aldehyde transfer	Beriberi (weight loss, heart problems, neurological dysfunction)
Riboflavin (B_2)	Flavin adenine dinucleotide (FAD)	Oxidation–reduction	Cheliosis and angular stomatitis (lesions of the mouth), dermatitis
Pyridoxine (B_6)	Pyridoxal phosphate	Group transfer to or from amino acids	Depression, confusion, convulsions
Nicotinic acid (niacin) (B_3)	Nicotinamide adenine dinucleotide (NAD^+)	Oxidation–reduction	Pellagra (dermatitis, depression, diarrhea)
Pantothenic acid (B_5)	Coenzyme A	Acyl-group transfer	Hypertension
Biotin (B_7)	Biotin–lysine adducts (biocytin)	ATP-dependent carboxylation and carboxyl-group transfer	Rash about the eyebrows, muscle pain, fatigue (rare)
Folic acid (B_9)	Tetrahydrofolate	Transfer of one-carbon components; thymine synthesis	Anemia, neural-tube defects in development
B_{12}	5′-Deoxyadenosyl cobalamin	Transfer of methyl groups; intramolecular rearrangements	Anemia, pernicious anemia, methylmalonic acidosis

activated carriers in all organisms is one of the unifying motifs of biochemistry. Furthermore, it illustrates the modular design of metabolism. A small set of molecules carries out a very wide range of tasks. Metabolism is readily comprehended because of the economy and elegance of its underlying design.

Many activated carriers are derived from vitamins

Almost all the activated carriers that act as coenzymes are derived from *vitamins*. Vitamins are organic molecules that are needed in small amounts in the diets of some higher animals. Table 15.3 lists the vitamins that act as coenzymes and Figure 15.17 shows the structures of some of them. This series of vitamins is known as the vitamin B group. In all cases, the vitamin must be modified before it can serve its function. We have already touched on the roles of niacin, riboflavin, and pantothenate. We will see these three and the other B vitamins many times in our study of biochemistry.

Vitamins serve the same roles in nearly all forms of life, but higher animals lost the capacity to synthesize them in the course of evolution. For instance, whereas *E. coli* can thrive on glucose and organic salts, human beings require at least 12 vitamins in their diet. The biosynthetic pathways for



Vitamin B_5 (Pantothenate)

Vitamin B_2 (Riboflavin)

Vitamin B_3 (Niacin)

Vitamin B_6 (Pyridoxine)

FIGURE 15.17 Structures of some of the B vitamins. These vitamins are often referred to as water-soluble vitamins because of the ease with which they dissolve in water.

TABLE 15.4 Noncoenzyme vitamins

Vitamin	Function	Deficiency
A	Roles in vision, growth, reproduction	Night blindness, cornea damage, damage to respiratory and gastrointestinal tract
C (ascorbic acid)	Antioxidant	Scurvy (swollen and bleeding gums, subdermal hemorrhaging)
D	Regulation of calcium and phosphate metabolism	Rickets (children): skeletal deformities, impaired growth Osteomalacia (adults): soft, bending bones
E	Antioxidant	Lesions in muscles and nerves (rare)
K	Blood coagulation	Subdermal hemorrhaging

vitamins can be complex; thus, it is biologically more efficient to ingest vitamins than to synthesize the enzymes required to construct them from simple molecules. This efficiency comes at the cost of dependence on other organisms for chemicals essential for life. Indeed, vitamin deficiency can generate diseases in all organisms requiring these molecules (Tables 15.3 and 15.4).

Not all vitamins function as coenzymes. Vitamins designated by the letters A, C, D, E, and K (Figure 15.18 and Table 15.4) have a diverse array of functions. Vitamin A (retinol) is the precursor of retinal, the light-sensitive group in rhodopsin and other visual pigments (Section 32.3), and retinoic acid, an important signaling molecule. A deficiency of this vitamin leads to night blindness. In addition, young animals require vitamin A for growth. Vitamin C, or ascorbate, acts as an antioxidant. A deficiency in vitamin C results in the formation of unstable collagen molecules and is the cause of scurvy, a disease characterized by skin lesions and blood-vessel fragility (Section 27.6). A metabolite of vitamin D is a hormone that regulates the metabolism of calcium and phosphorus. A deficiency in vitamin D impairs bone formation in growing animals. Vitamin E (α-tocopherol) deficiency causes a variety of neuromuscular pathologies. This vitamin inactivates reactive oxygen species such as hydroxyl

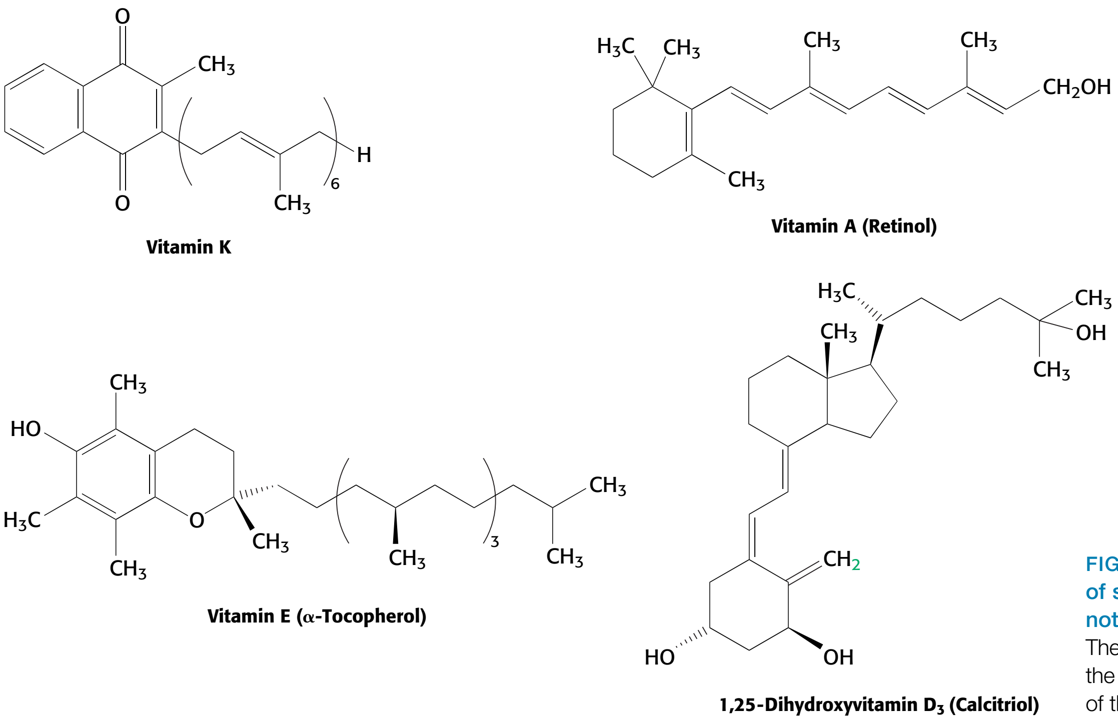


FIGURE 15.18 Structures of some vitamins that do not function as coenzymes. These vitamins are often called the fat-soluble vitamins because of their hydrophobic nature.

TABLE 15.5 Types of chemical reactions in metabolism

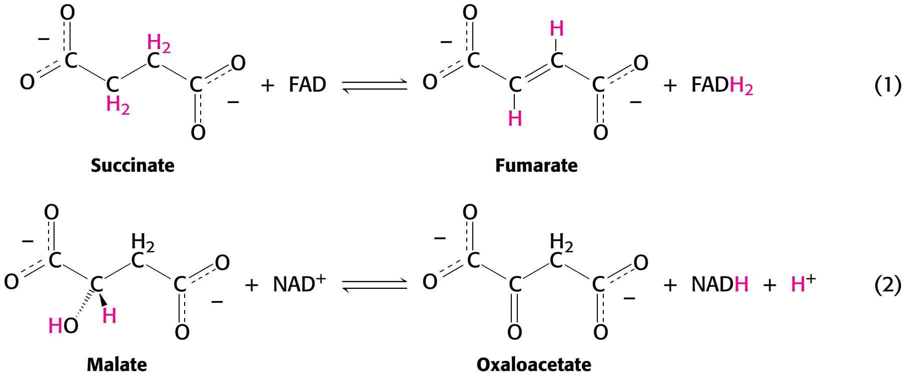
Type of reaction	Description
Oxidation–reduction	Electron transfer
Ligation requiring ATP cleavage	Formation of covalent bonds (i.e., carbon–carbon bonds)
Isomerization	Rearrangement of atoms to form isomers
Group transfer	Transfer of a functional group from one molecule to another
Hydrolytic	Cleavage of bonds by the addition of water
Carbon bond cleavage by means other than hydrolysis or oxidation	Two substrates yielding one product or vice versa. When H_2O or CO_2 are a product, a double bond is formed.

radicals before they can oxidize unsaturated membrane lipids, damaging cell structures. Vitamin K is required for normal blood clotting (Section 10.4).

Key reactions are reiterated throughout metabolism

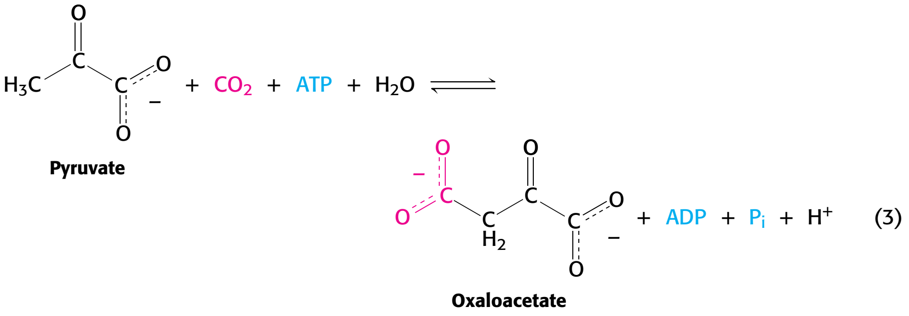
Just as there is an economy of design in the use of activated carriers, so is there an economy of design in biochemical reactions. The thousands of metabolic reactions, bewildering at first in their variety, can be subdivided into just six types (Table 15.5). Specific reactions of each type appear repeatedly, reducing the number of reactions that a student needs to learn.

1. *Oxidation–reduction reactions* are essential components of many pathways. Useful energy is often derived from the oxidation of carbon compounds. Consider the following two reactions:



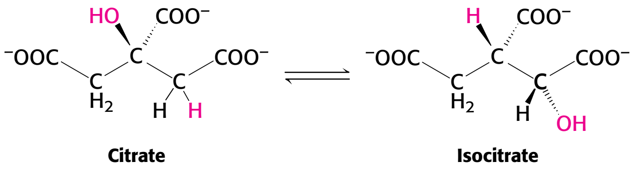
These two oxidation–reduction reactions are components of the citric acid cycle (Chapter 17), which completely oxidizes the activated two-carbon fragment of acetyl CoA to two molecules of CO_2. In reaction 1, $FADH_2$ carries the electrons, whereas, in reaction 2, electrons are carried by NADH.

2. *Ligation reactions* form bonds by using free energy from ATP cleavage. Reaction 3 illustrates the ATP-dependent formation of a carbon–carbon bond, necessary to combine smaller molecules to form larger ones. Oxaloacetate is formed from pyruvate and CO_2.



The oxaloacetate can be used in the citric acid cycle, or converted into glucose or amino acids such as aspartic acid.

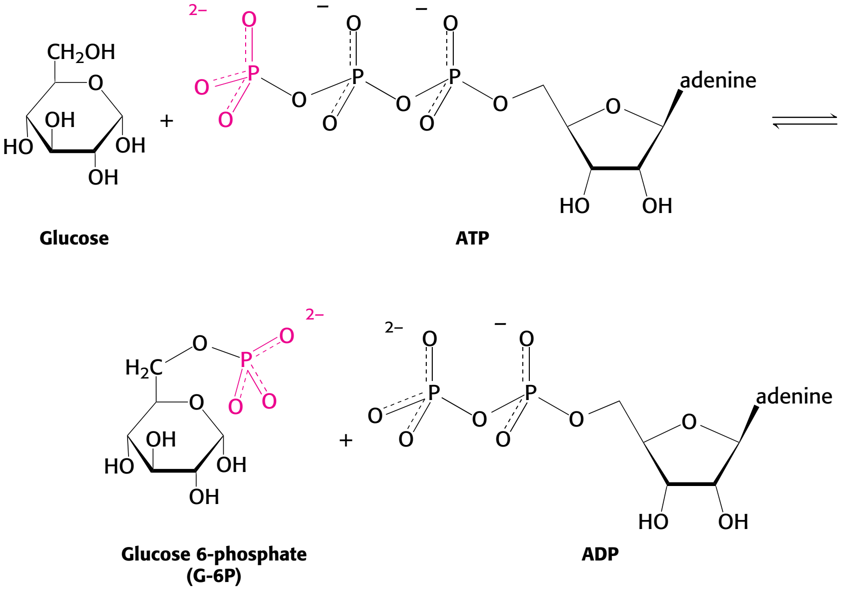
3. *Isomerization reactions* rearrange particular atoms within a molecule. Their role is often to prepare the molecule for subsequent reactions such as the oxidation–reduction reactions described in point 1.


HO
COO⁻
⁻OOC
C
COO⁻
C
H₂
C
H
H
Citrate
H
COO⁻
⁻OOC
C
COO⁻
C
H₂
C
H
OH
Isocitrate

(4)

Reaction 4 is, again, a component of the citric acid cycle. This isomerization prepares the molecule for subsequent oxidation and decarboxylation reactions by moving the hydroxyl group of citrate from a tertiary to a secondary position.

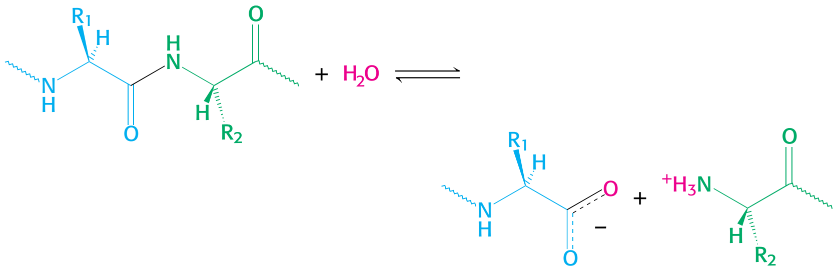
4. *Group-transfer reactions* play a variety of roles. Reaction 5 is representative of such a reaction. A phosphoryl group is transferred from the activated phosphoryl-group carrier, ATP, to glucose, the initial step in glycolysis, a key pathway for extracting energy from glucose (Chapter 16). This reaction traps glucose in the cell so that further catabolism can take place.


CH₂OH
O
OH
HO
OH
OH
Glucose
+
2−
O
P
O
O
O
−
O
P
O
O
−
O
P
O
O
O
adenine
O
HO
OH
ATP
H₂C
O
P
O
O
O
2−
O
OH
HO
OH
OH
Glucose 6-phosphate
(G-6P)
+
2−
O
P
O
O
O
−
O
P
O
O
adenine
O
HO
OH
ADP

(5)

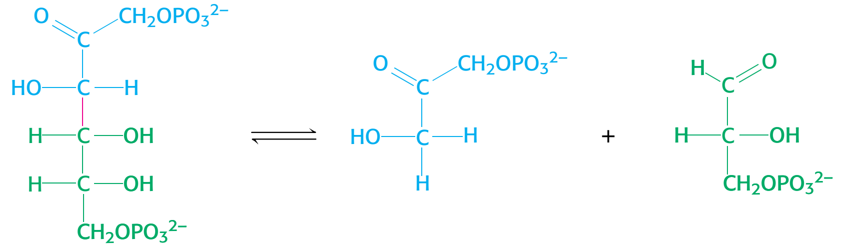
As stated earlier, group-transfer reactions are used to synthesize ATP. We also saw examples of their use in signaling pathways (Chapter 14).

5. *Hydrolytic reactions* cleave bonds by the addition of water. Hydrolysis is a common means of degrading large molecules, either to facilitate further metabolism or to reuse some of the components for biosynthetic purposes. Proteins are digested by hydrolytic cleavage (chapters 9 and 10). Reaction 6 illustrates the hydrolysis of a peptide to yield two smaller peptides.



(6)

6. *Carbon bonds can be cleaved by means other than hydrolysis or oxidation, with two substrates yielding one product or vice versa.* When CO_2 or H_2O is released, a double bond is formed. The enzymes that catalyze these types of reaction are classified as *lyases.* An important example, illustrated in reaction 7, is the conversion of the six-carbon molecule fructose 1,6-bisphosphate into two three-carbon fragments: dihydroxyacetone phosphate and glyceraldehyde 3-phosphate.



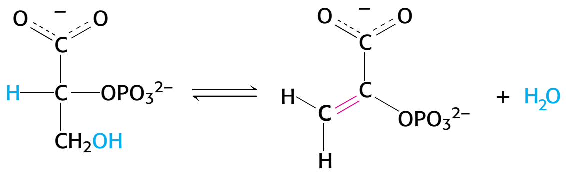
(7)

Fructose 1,6-bisphosphate (F-1,6-BP)

Dihydroxyacetone phosphate (DHAP)

Glyceraldehyde 3-phosphate (GAP)

This reaction is a critical step in glycolysis (Chapter 16). Dehydrations to form double bonds, such as the formation of phosphoenolpyruvate (Figure 15.6) from 2-phosphoglycerate (reaction 8), are important reactions of this type.



(8)

2-Phosphoglycerate

Phosphoenolpyruvate (PEP)

The dehydration sets up the next step in the pathway, a group-transfer reaction that uses the high phosphoryl-transfer potential of the product PEP to form ATP from ADP.

These six fundamental reaction types are the basis of metabolism. Remember that all six types can proceed in either direction, depending on the standard free energy for the specific reaction and the concentrations of the reactants and products inside the cell. An effective way to learn is to look for commonalities in the diverse metabolic pathways that we will be examining. There is a chemical logic that, when exposed, renders the complexity of the chemistry of living systems more manageable and reveals its elegance.

Metabolic processes are regulated in three principal ways

It is evident that the complex network of metabolic reactions must be rigorously regulated. The levels of available nutrients must be monitored and the activity of metabolic pathways must be altered and integrated to create *homeostasis*, a stable biochemical environment. At the same time, metabolic control must be flexible, able to adjust metabolic activity to the constantly changing external environments of cells. Figure 15.19 illustrates the nutrient

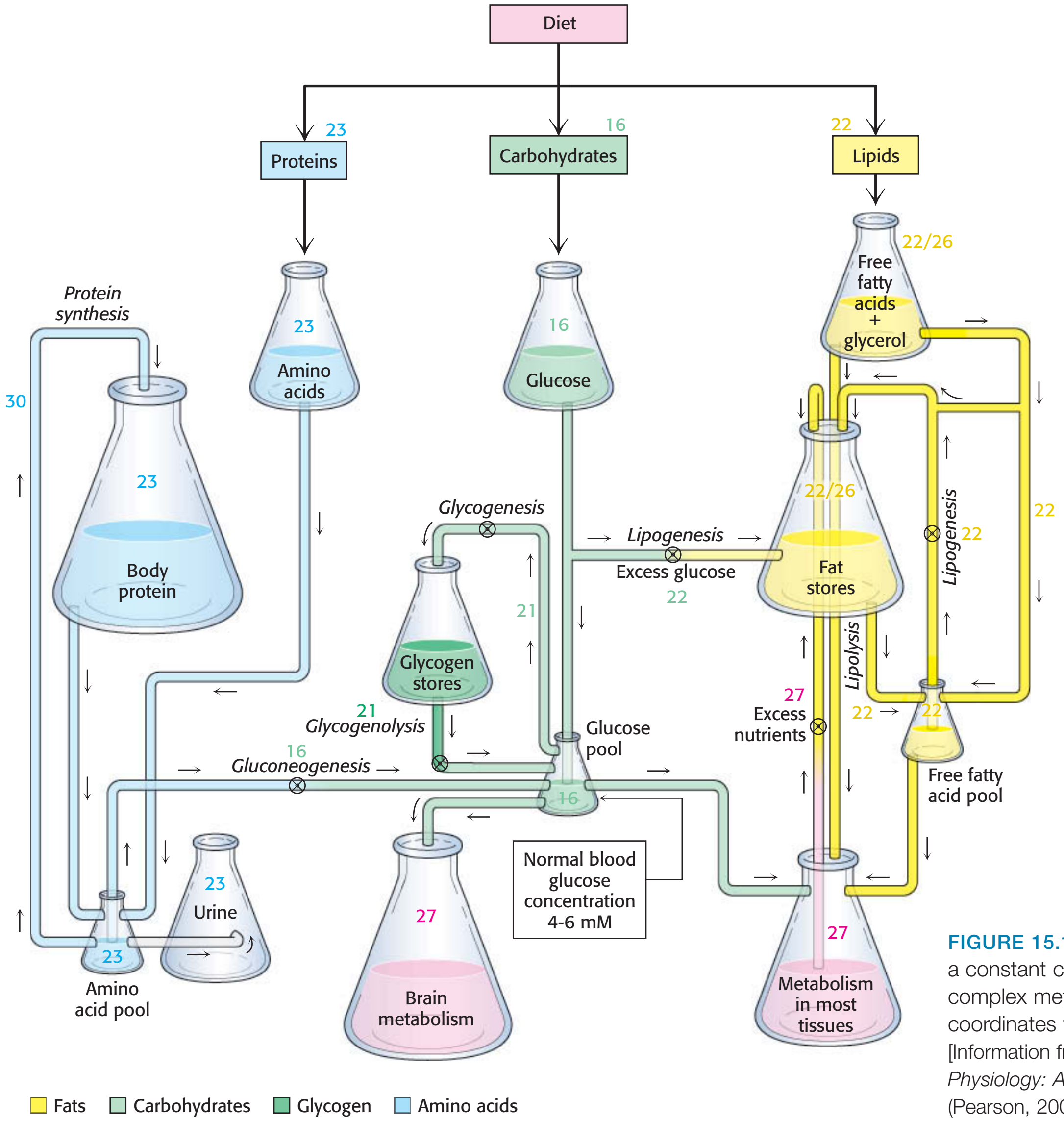

Diet
23
Proteins
16
Carbohydrates
22
Lipids
22/26
Free fatty acids + glycerol
Protein synthesis
30
23
Amino acids
16
Glucose
23
Body protein
Glycogenesis
Lipogenesis
Excess glucose
22
22/26
Fat stores
Lipogenesis
22
22
21
Glycogen stores
21
Glycogenolysis
16
Gluconeogenesis
Glucose pool
16
27
Excess nutrients
Lipolysis
22
22
Free fatty acid pool
Normal blood glucose concentration 4-6 mM
23
Urine
23
Amino acid pool
27
Brain metabolism
27
Metabolism in most tissues
🟨 Fats
🟩 Carbohydrates
🟩 Glycogen
🟦 Amino acids

FIGURE 15.19 Homeostasis. Maintaining a constant cellular environment requires complex metabolic regulation that coordinates the use of nutrient pools. [Information from D. U. Silverthorn, *Human Physiology: An Integrated Approach*, 3rd ed. (Pearson, 2004), Figure 22-2.]

pools and their connections that must be monitored and regulated. Metabolism is regulated through control of (1) *the amounts of enzymes,* (2) *their catalytic activities,* and (3) *the accessibility of substrates.*

Controlling the amounts of enzymes. The amount of a particular enzyme depends on both its rate of synthesis and its rate of degradation. The level of many enzymes is adjusted by a change in the *rate of transcription* of the genes encoding them (chapters 29 and 31). In *E. coli,* for example, the presence of lactose induces within minutes a more than 50-fold increase in the rate of synthesis of β-galactosidase, the enzyme required for the breakdown of this disaccharide.

Controlling catalytic activity. The catalytic activity of enzymes is controlled in several ways. *Allosteric control* is especially important. For example, the first reaction in many biosynthetic pathways is allosterically inhibited by the ultimate product of the pathway. The inhibition of aspartate transcarbamoylase by cytidine triphosphate (Section 10.1) is a well-understood example of *feedback inhibition*. This type of control can be almost instantaneous. Another recurring mechanism is *reversible covalent modification* (Section 10.3). For example, glycogen phosphorylase, the enzyme catalyzing the breakdown of glycogen, a storage form of sugar, is activated by the phosphorylation of a particular serine residue when glucose is scarce (Section 21.1).

Hormones coordinate metabolic relations between different tissues, often by regulating the reversible modification of key enzymes. For instance, the hormone epinephrine triggers a signal-transduction cascade in muscle, resulting in the phosphorylation and activation of key enzymes and leading to the rapid degradation of glycogen to glucose, which is then used to supply ATP for muscle contraction. As described in Chapter 14, many hormones act through intracellular messengers, such as cyclic AMP and calcium ion, that coordinate the activities of many target proteins.

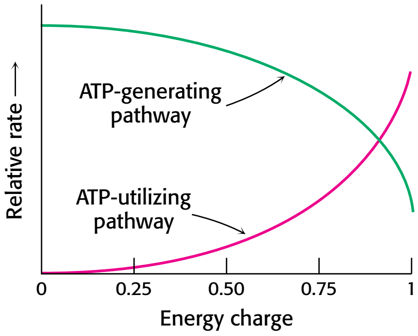


FIGURE 15.20 Energy charge regulates metabolism. When the energy charge is high, ATP inhibits the relative rates of a typical ATP-generating (catabolic) pathway and stimulates the typical ATP-utilizing (anabolic) pathway.

Many reactions in metabolism are controlled by the *energy status* of the cell. One index of the energy status is the *energy charge,* which is proportional to the mole fraction of ATP plus half the mole fraction of ADP, given that ATP contains two anhydride bonds, whereas ADP contains one. Hence, the energy charge is defined as

$$\text{Energy charge} = \frac{[\text{ATP}] + \frac{1}{2}[\text{ADP}]}{[\text{ATP}] + [\text{ADP}] + [\text{AMP}]}$$

The energy charge can have a value ranging from 0 (all AMP) to 1 (all ATP). *ATP-generating (catabolic) pathways are inhibited by a high energy charge, whereas ATP-utilizing (anabolic) pathways are stimulated by a high energy charge.* In plots of the reaction rates of such pathways versus the energy charge, the curves are steep near an energy charge of 0.9, where they usually intersect (Figure 15.20). It is evident that the control of these pathways has evolved to maintain the energy charge within rather narrow limits. In other words, *the energy charge, like the pH of a cell, is buffered.* The energy charge of most cells ranges from 0.90 to 0.95, but can fall to less than 0.7 in muscle during high-intensity exercise. An alternative index of the energy status is the *phosphorylation potential,* which is defined as

$$\text{Phosphorylation potential} = \frac{[\text{ATP}]}{[\text{ADP}] + [\text{P}_i]}$$

The phosphorylation potential, in contrast with the energy charge, depends on the concentration of P_i and is directly related to the free-energy storage available from ATP.

Controlling the accessibility of substrates. Controlling the *availability of substrates* is another means of regulating metabolism in all organisms. For instance, glucose breakdown can take place in many cells only if insulin is present to promote the entry of glucose into the cell. In eukaryotes, metabolic regulation and flexibility are enhanced by compartmentalization. The transfer of substrates from one compartment of a cell to another can serve as a control point. For example, fatty acid oxidation takes place in mitochondria, whereas fatty acid synthesis takes place in the cytoplasm. *Compartmentalization segregates opposed reactions.*

Aspects of metabolism may have evolved from an RNA world

How did the complex pathways that constitute metabolism evolve? The current thinking is that RNA was an early biomolecule that dominated metabolism, serving both as a catalyst and an information storage molecule. This hypothetical time is called the RNA world.

Why do activated carriers such as ATP, NADH, $FADH_2$, and coenzyme A contain adenosine diphosphate units? A possible explanation is that these molecules evolved from the early RNA catalysts. Non-RNA units such as the isoalloxazine ring may have been recruited to serve as efficient carriers of activated electrons and chemical units, a function not readily performed by RNA itself. We can picture the adenine ring of $FADH_2$ binding to a uracil unit in a niche of an RNA enzyme (ribozyme) by base-pairing, whereas the isoalloxazine ring protrudes and functions as an electron carrier. When the more versatile proteins replaced RNA as the major catalysts, the ribonucleotide coenzymes stayed essentially unchanged because they were already well suited to their metabolic roles. The nicotinamide unit of NADH, for example, can readily transfer electrons irrespective of whether the adenine unit interacts with a base in an RNA enzyme or with amino acid residues in a protein enzyme. With the advent of protein enzymes, these important cofactors evolved as free molecules without losing the adenosine diphosphate vestige of their RNA-world ancestry. That molecules and motifs of metabolism are common to all forms of life testifies to their common origin and to the retention of functioning modules through billions of years of evolution. Our understanding of metabolism, like that of other biological processes, is enriched by inquiry into how these beautifully integrated patterns of reactions came into being.

SUMMARY

All cells transform energy. They extract energy from their environment and use this energy to convert simple molecules into cellular components.

15.1 Metabolism Is Composed of Many Coupled, Interconnecting Reactions

The process of energy transduction takes place through metabolism, a highly integrated network of chemical reactions. Metabolism can be subdivided into catabolism (reactions employed to extract energy from fuels) and anabolism (reactions that use this energy for biosynthesis). The most valuable thermodynamic concept for understanding bioenergetics is free energy. A reaction can occur spontaneously only if the change in free energy (ΔG) is negative. A thermodynamically unfavorable reaction can be driven by a thermodynamically favorable one, which is the hydrolysis of ATP in many cases.

15.2 ATP Is the Universal Currency of Free Energy in Biological Systems

The energy derived from catabolism is transformed into adenosine triphosphate. ATP hydrolysis is exergonic and the energy released can be used to power cellular processes, including motion, active transport, and biosynthesis. Under cellular conditions, the hydrolysis of ATP shifts the equilibrium of a coupled reaction by a factor of 10^8. ATP, the universal currency of energy in biological systems, is an energy-rich molecule because it contains two phosphoanhydride bonds.

15.3 The Oxidation of Carbon Fuels Is an Important Source of Cellular Energy

ATP formation is coupled to the oxidation of carbon fuels, either directly or through the formation of ion gradients. Photosynthetic organisms can use light to generate such gradients. ATP is consumed to drive endergonic reactions and in signal-transduction processes. The extraction of energy from foodstuffs by aerobic organisms comprises three stages. In the first stage, large molecules are broken down into smaller ones, such as amino acids, sugars, and fatty acids. In the second stage, these small molecules are degraded to a few simple units, such as acetyl CoA, that have widespread roles in metabolism. The third stage of metabolism is the citric acid cycle and oxidative phosphorylation, in which ATP is generated as electrons flow to O_2, the ultimate electron acceptor, and fuels are completely oxidized to CO_2.

15.4 Metabolic Pathways Contain Many Recurring Motifs

Metabolism is characterized by common motifs. A small number of recurring activated carriers, such as ATP, NADH, and acetyl CoA, transfer activated groups in many metabolic pathways. NADPH, which carries two electrons at a high potential, provides reducing power for reductive biosynthesis of cell components. Many activated carriers are derived from vitamins, small organic molecules required in the diets of many higher organisms. Moreover, key reaction types are used repeatedly in metabolic pathways.

Metabolism is regulated in a variety of ways. The amounts of some critical enzymes are controlled by regulation of the rate of synthesis and degradation. In addition, the catalytic activities of many enzymes are regulated by allosteric interactions and by covalent modification. The movement of many substrates into cells and subcellular compartments also is controlled. The energy charge, which depends on the relative amounts of ATP, ADP, and AMP, plays a role in metabolic regulation. A high energy charge inhibits ATP-generating (catabolic) pathways, whereas it stimulates ATP-utilizing (anabolic) pathways.

KEY TERMS

metabolism or intermediary metabolism (p. 423)
phototroph (p. 424)
chemotroph (p. 424)
catabolism (p. 424)
anabolism (p. 424)
amphibolic pathway (p. 425)
adenosine triphosphate (ATP) (p. 426)
phosphoryl-transfer potential (p. 429)
activated carrier (p. 435)
vitamin (p. 438)
oxidation–reduction reaction (p. 440)
ligation reaction (p. 440)
isomerization reaction (p. 441)
group-transfer reaction (p. 441)
hydrolytic reaction (p. 441)
cleavage of carbon bonds by means other than hydrolysis or oxidation (p. 442)
lyase (p. 442)
energy charge (p. 444)
phosphorylation potential (p. 444)

PROBLEMS

1. *Complex patterns.* What is meant by *intermediary metabolism?*

2. *Opposites.* Differentiate between anabolism and catabolism.

3. *Graffiti.* While walking to biochemistry class with a friend, you see the following graffiti spray painted on the wall of the science building: "When a system is in equilibrium, the Gibbs free energy is maximum." You are disgusted, not only at the vandalism, but at the ignorance of the vandal. Your friend asks you to explain.

4. *Why bother to eat?* What are the three primary uses for cellular energy?

5. *Match 'em.*

1. Cellular energy currency ____
2. Anabolic electron carrier ____
3. Phototroph ____
4. Catabolic electron carrier reaction ____
5. Oxidation-reduction reaction ____
6. Activated carrier of two carbon fragments ____
7. Vitamin ____
8. Anabolism ____
9. Amphibolic reaction ____
10. Catabolism ____

a. NAD^+
b. Coenzyme A
c. Precursor to coenzymes
d. Yields energy
e. Requires energy
f. ATP
g. Transfers electrons
h. $NADP^+$
i. Converts light energy to chemical energy
j. Used in anabolism and catabolism

6. *Charges.* In vivo, ATP is usually bound to magnesium or manganese ions. Why is this the case?

7. *Energy to burn.* What factors account for the high phosphoryl-transfer potential of nucleoside triphosphates?

8. *Back in time.* Account for the fact that ATP, and not another nucleoside triphosphate, is the cellular energy currency.

9. *Currency issues.* Why does it make good sense to have a single nucleotide, ATP, function as the cellular energy currency?

10. *Environmental conditions.* The standard free energy of hydrolysis for ATP is $-30.5\ \text{kJ mol}^{-1}$ ($-7.3\ \text{kcal mol}^{-1}$).

$$\text{ATP} + \text{H}_2\text{O} \rightleftharpoons \text{ADP} + \text{P}_\text{i}$$

What conditions might be changed to alter the free energy of hydrolysis?

11. *Brute force?* Metabolic pathways frequently contain reactions with positive standard free energy values, yet the reactions still take place. How is this possible?

12. *Energy flow.* What is the direction of each of the following reactions when the reactants are initially present in equimolar amounts? Use the data given in Table 15.1.

(a) ATP + creatine $\rightleftharpoons$ creatine phosphate + ADP

(b) ATP + glycerol $\rightleftharpoons$ glycerol 3-phosphate + ADP

(c) ATP + pyruvate $\rightleftharpoons$ phosphoenolpyruvate + ADP

(d) ATP + glucose $\rightleftharpoons$ glucose 6-phosphate + ADP

13. *A proper inference.* What information do the $\Delta G^{\circ\prime}$ data given in Table 15.1 provide about the relative rates of hydrolysis of pyrophosphate and acetyl phosphate?

14. *A potent donor.* Consider the following reaction:

ATP + pyruvate $\rightleftharpoons$ phosphoenolpyruvate + ADP

(a) Calculate $\Delta G^{\circ\prime}$ and K'_{eq} at 25°C for this reaction by using the data given in Table 15.1.

(b) What is the equilibrium ratio of pyruvate to phosphoenolpyruvate if the ratio of ATP to ADP is 10?

15. *Isomeric equilibrium.* Using the information in Table 15.1, calculate $\Delta G^{\circ\prime}$ for the isomerization of glucose 6-phosphate to glucose 1-phosphate. What is the equilibrium ratio of glucose 6-phosphate to glucose 1-phosphate at 25°C?

16. *Activated acetate.* The formation of acetyl CoA from acetate is an ATP-driven reaction:

Acetate + ATP + CoA $\rightleftharpoons$ acetyl CoA + AMP + PP_i

(a) Calculate $\Delta G^{\circ\prime}$ for this reaction by using data given in this chapter.

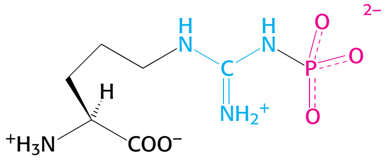
(b) The PP_i formed in the preceding reaction is rapidly hydrolyzed in vivo because of the ubiquity of inorganic pyrophosphatase. The $\Delta G^{\circ\prime}$ for the hydrolysis of PP_i is -19.2 kJ mol^{-1} (-4.6 kcal mol^{-1}). Calculate the $\Delta G^{\circ\prime}$ for the overall reaction, including pyrophosphate hydrolysis. What effect does the hydrolysis of PP_i have on the formation of acetyl CoA?

17. *Acid strength.* The p*K* of an acid is a measure of its proton-group-transfer potential.

(a) Derive a relation between $\Delta G^{\circ\prime}$ and p*K*.

(b) What is the $\Delta G^{\circ\prime}$ for the ionization of acetic acid, which has a p*K* of 4.8?

18. *Raison d'être.* The muscles of some invertebrates are rich in arginine phosphate (phosphoarginine). Propose a function for this amino acid derivative.



Arginine phosphate

19. *Recurring motif.* What is the structural feature common to ATP, FAD, NAD^+, and CoA?

20. *Ergogenic help or hindrance?* Creatine is a popular, but untested, dietary supplement.

(a) What is the biochemical rationale for the use of creatine?

(b) What type of exercise would most benefit from creatine supplementation?

21. *Standard conditions versus real life 1.* The enzyme aldolase catalyzes the following reaction in the glycolytic pathway:

$$\text{Fructose 1, 6-bisphosphate} \xrightleftharpoons{\text{Aldolase}} \text{dihydroxyacetone phosphate} + \text{glyceraldehyde 3-phosphate}$$

The $\Delta G^{\circ\prime}$ for the reaction is $+23.8$ kJ mol^{-1} ($+5.7$ kcal mol^{-1}), whereas the ΔG in the cell is -1.3 kJ mol^{-1} (-0.3 kcal mol^{-1}). Calculate the ratio of reactants to products under equilibrium and intracellular conditions. Using your results, explain how the reaction can be endergonic under standard conditions and exergonic under intracellular conditions.

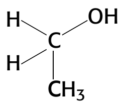
22. *Standard conditions versus real life 2.* On page 428, we showed that a reaction, A $\rightleftharpoons$ B, with a $\Delta G^{\circ\prime} = +16.7$ kJ mol^{-1} ($+4.0$ kcal mol^{-1}) has an K'_{eq} of 1.15×10^{-3}. The K'_{eq} is increased to 2.67×10^{2} if the reaction is coupled to ATP hydrolysis under standard conditions. The ATP-generating system of cells maintains the $[ATP]/[ADP][P_i]$ ratio at a high level, typically of the order of 500 M^{-1}. Calculate the ratio of B/A under cellular conditions.

23. *Not all alike.* The concentrations of ATP, ADP, and P_i differ with cell type. Consequently, the release of free energy with the hydrolysis of ATP will vary with cell type. Using the following table, calculate the ΔG for the hydrolysis of ATP in liver, muscle, and brain cells. In which cell type is the free energy of ATP hydrolysis most negative?

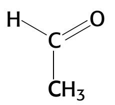
	ATP (mM)	ADP (mM)	P_i (mM)
Liver	3.5	1.8	5.0
Muscle	8.0	0.9	8.0
Brain	2.6	0.7	2.7

24. *Oxidation issues.* Examine the pairs of molecules and identify the more-reduced molecule in each pair.

(a)

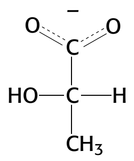


Ethanol

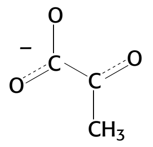


Acetaldehyde

(b)

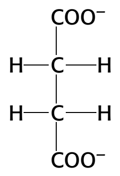


Lactate

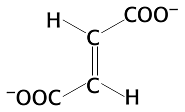


Pyruvate

(c)

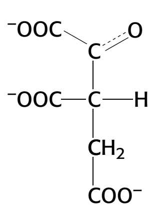


Succinate

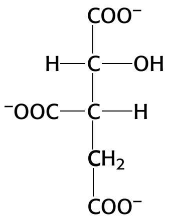


Fumarate

(d)

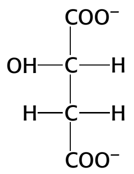


Oxalosuccinate

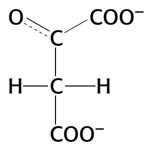


Isocitrate

(e)



Malate



Oxaloacetate

25. *Running downhill*. Glycolysis is a series of 10 linked reactions that convert one molecule of glucose into two molecules of pyruvate with the concomitant synthesis of two molecules of ATP (Chapter 16). The $\Delta G^{\circ\prime}$ for this set of reactions is -35.6 kJ mol^{-1} (-8.5 kcal mol^{-1}), whereas the ΔG is -90 kJ mol^{-1} (-22 kcal mol^{-1}). Explain why the free-energy release is so much greater under intracellular conditions than under standard conditions.

26. *Outsourcing*. Outsourcing, a common business practice, is contracting with another business to perform a particular function. Higher organisms were the original outsourcers, frequently depending on lower organisms to perform key biochemical functions. Give an example from this chapter of biochemical outsourcing.

27. *Breakdown products*. Digestion is the first stage in the extraction of energy from food, but no useful energy is acquired during this stage. Why is digestion considered a stage in energy extraction?

28. *High-energy electrons*. What are the activated electron carriers for catabolism? For anabolism?

29. *Less reverberation*. Thioesters, common in biochemistry, are more unstable (energy-rich) than oxygen esters. Explain why this is the case.

30. *Classifying reactions*. What are the six common types of reactions seen in biochemistry?

31. *Staying in control*. What are the three principal means of controlling metabolic reactions?

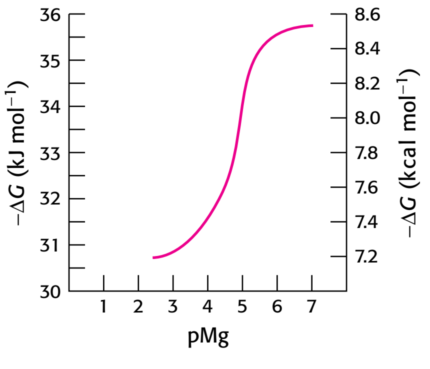
Chapter Integration Problems

32. *Kinetic versus thermodynamic*. The reaction of NADH with oxygen to produce NAD^+ and H_2O is very exergonic, yet the reaction of NADH and oxygen takes place very slowly. Why does a thermodynamically favorable reaction not occur rapidly?

33. *Activated sulfate*. Fibrinogen contains tyrosine-*O*-sulfate. Propose an activated form of sulfate that could react in vivo with the aromatic hydroxyl group of a tyrosine residue in a protein to form tyrosine-*O*-sulfate.

Data Interpretation Problem

34. *Opposites attract*. The following graph shows how the ΔG for the hydrolysis of ATP varies as a function of the Mg^{2+} concentration (pMg $= -\log[Mg^{2+}]$).



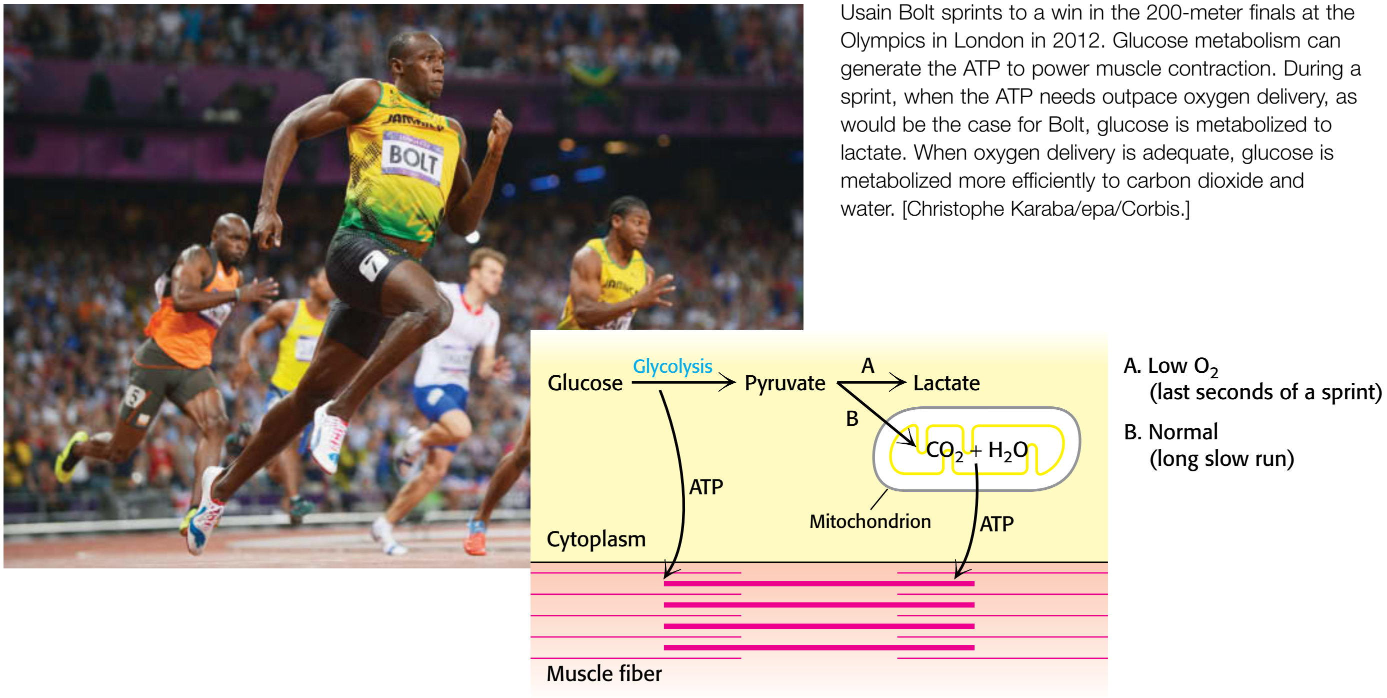
(a) How does decreasing $[Mg^{2+}]$ affect the ΔG of hydrolysis for ATP?

(b) Explain this effect.

CHAPTER

16

Glycolysis and Gluconeogenesis



Usain Bolt sprints to a win in the 200-meter finals at the Olympics in London in 2012. Glucose metabolism can generate the ATP to power muscle contraction. During a sprint, when the ATP needs outpace oxygen delivery, as would be the case for Bolt, glucose is metabolized to lactate. When oxygen delivery is adequate, glucose is metabolized more efficiently to carbon dioxide and water. [Christophe Karaba/epa/Corbis.]

The first metabolic pathway that we encounter is *glycolysis,* an ancient pathway employed by a host of organisms. *Glycolysis is the sequence of reactions that metabolizes one molecule of glucose to two molecules of pyruvate with the concomitant net production of two molecules of ATP.* This process is anaerobic (i.e., it does not require O_2) because it evolved before substantial amounts of oxygen accumulated in the atmosphere. Pyruvate can be further processed anaerobically to lactate (*lactic acid fermentation*) or ethanol (*alcoholic fermentation*). Under aerobic conditions, pyruvate can be completely oxidized to CO_2, generating much more ATP, as will be described in chapters 17 and 18. Figure 16.1 shows some possible fates of pyruvate produced by glycolysis.

Because glucose is such a precious fuel, metabolic products, such as pyruvate and lactate, are salvaged to synthesize glucose in the process of *gluconeogenesis.* Although glycolysis and gluconeogenesis have some enzymes in common, the two pathways are not simply the reverse of each other. In particular, the highly exergonic, irreversible steps of glycolysis are bypassed in gluconeogenesis. The two pathways are reciprocally regulated so that glycolysis and gluconeogenesis do not take place simultaneously in the same cell to a significant extent.

OUTLINE

16.1 Glycolysis Is an Energy-Conversion Pathway in Many Organisms
16.2 The Glycolytic Pathway Is Tightly Controlled
16.3 Glucose Can Be Synthesized from Noncarbohydrate Precursors
16.4 Gluconeogenesis and Glycolysis Are Reciprocally Regulated

Glycolysis
Derived from the Greek stem *glyk-*, “sweet,” and the word *lysis*, “dissolution.”

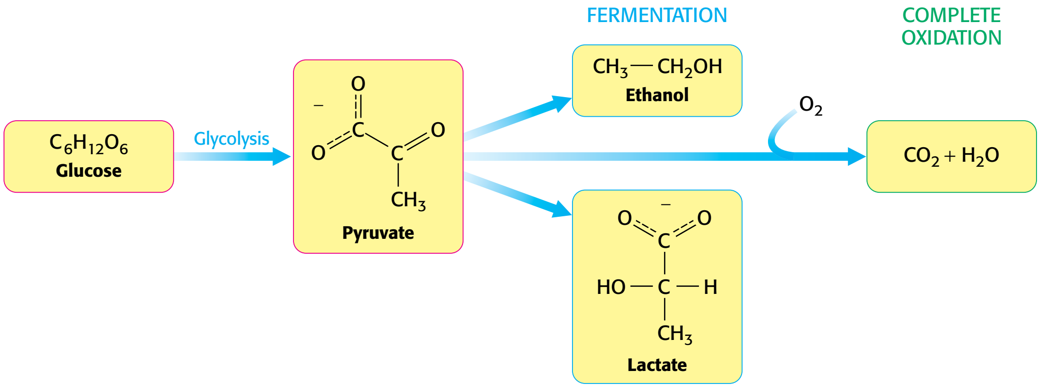


FIGURE 16.1 Some fates of glucose.

Our understanding of glucose metabolism, especially glycolysis, has a rich history. Indeed, the development of biochemistry and the delineation of glycolysis went hand in hand. A key discovery was made by Hans and Eduard Buchner in 1897, quite by accident. The Buchners were interested in manufacturing cell-free extracts of yeast for possible therapeutic use. These extracts had to be preserved without the use of antiseptics such as phenol, and so they decided to try sucrose, a commonly used preservative in kitchen chemistry. They obtained a startling result: sucrose was rapidly fermented into alcohol by the yeast juice. The significance of this finding was immense. *The Buchners demonstrated for the first time that fermentation could take place outside living cells.* The accepted view of their day, asserted by Louis Pasteur in 1860, was that fermentation is inextricably tied to living cells. The chance discovery by the Buchners refuted this dogma and opened the door to modern biochemistry. The Buchners' discovery inspired the search for the biochemicals that catalyze the conversion of sucrose into alcohol. *The study of metabolism became the study of chemistry.*

Studies of muscle extracts then showed that many of the reactions of lactic acid fermentation were the same as those of alcoholic fermentation. *This exciting discovery revealed an underlying unity in biochemistry.* The complete glycolytic pathway was elucidated by 1940. Glycolysis is also known as the *Embden–Meyerhof pathway*, after two pioneers of research on glycolysis.

Enzyme

A term coined by Friedrich Wilhelm Kühne in 1878 to designate catalytically active substances that had formerly been called ferments. Derived from the Greek words *en*, "in," and *zyme*, "leaven."

Glucose is generated from dietary carbohydrates

We typically consume in our diets a generous amount of starch and a smaller amount of glycogen. These complex carbohydrates must be converted into simpler carbohydrates for absorption by the intestine and transport in the blood. Starch and glycogen are digested primarily by the pancreatic enzyme *α-amylase* and to a lesser extent by salivary α-amylase. Amylase cleaves the α-1,4 bonds of starch and glycogen, but not the α-1,6 bonds. The products are the di- and trisaccharides maltose and maltotriose. The material not digestible because of the α-1,6 bonds is called the *limit dextrin.*

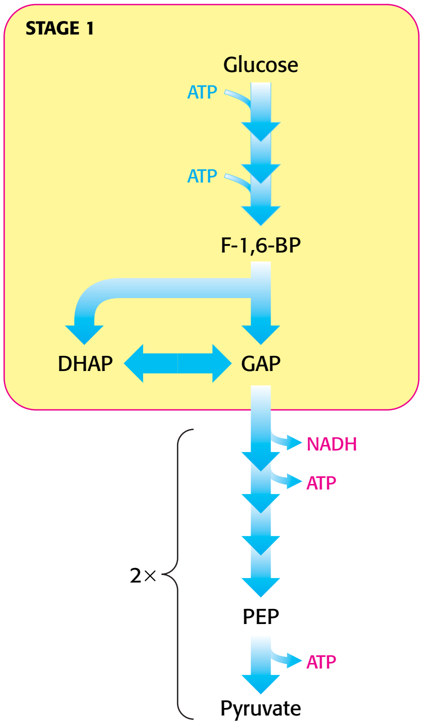
Maltase cleaves maltose into two glucose molecules, whereas *α-glucosidase* digests maltotriose and any other oligosaccharides that may have escaped digestion by the amylase. *α-Dextrinase* further digests the limit dextrin. Maltase and α-glucosidase are located on the surface of the intestinal cells, as is *sucrase,* an enzyme that degrades the sucrose contributed by vegetables to fructose and glucose. The enzyme *lactase* is responsible for degrading the milk sugar lactose into glucose and galactose. The monosaccharides are transported into the cells lining the intestine and then into the bloodstream.

Glucose is an important fuel for most organisms

Glucose is a common and important fuel. In mammals, glucose is the only fuel that the brain uses under nonstarvation conditions and the only fuel that red blood cells can use at all. Indeed, almost all organisms use glucose, and most that do process it in a similar fashion. Recall from Chapter 11 that there are many carbohydrates. Why is glucose instead of some other monosaccharide such a prominent fuel? We can speculate on the reasons. First, glucose is one of several monosaccharides formed from formaldehyde under prebiotic conditions, and so it may have been available as a fuel source for primitive biochemical systems. Second, glucose has a low tendency, relative to other monosaccharides, to nonenzymatically glycosylate proteins. In their open-chain forms, monosaccharides contain carbonyl groups that can react with the amino groups of proteins to form Schiff bases, which rearrange to form a more stable amino–ketone linkage (p. 321). Such nonspecifically modified proteins often do not function effectively. Glucose has a strong tendency to exist in the ring conformation and, consequently, relatively little tendency to modify proteins. Recall that all the hydroxyl groups in the ring conformation of β-glucose are equatorial, contributing to the sugar's high relative stability (p. 320).

16.1 Glycolysis Is an Energy-Conversion Pathway in Many Organisms

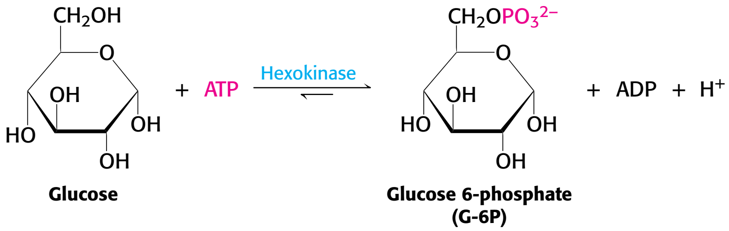
We now begin our consideration of the glycolytic pathway. This pathway is common to virtually all cells, both prokaryotic and eukaryotic. In eukaryotic cells, glycolysis takes place in the cytoplasm. This pathway can be thought of as comprising two stages (Figure 16.2). Stage 1 is the trapping and preparation phase. No ATP is generated in this stage. In stage 1, glucose is converted into fructose 1,6-bisphosphate in three steps: a phosphorylation, an isomerization, and a second phosphorylation reaction. *The strategy of these initial steps in glycolysis is to trap the glucose in the cell and form a compound that can be readily cleaved into phosphorylated three-carbon units.* Stage 1 is completed with the cleavage of the fructose 1,6-bisphosphate into two three-carbon fragments. These resulting three-carbon units are readily interconvertible. In stage 2, ATP is harvested when the three-carbon fragments are oxidized to pyruvate.



First stage of glycolysis. The first stage of glycolysis begins with the phosphorylation of glucose by hexokinase and ends with the isomerization of dihydroxyacetone phosphate to glyceraldehyde 3-phosphate.

Hexokinase traps glucose in the cell and begins glycolysis

Glucose enters cells through specific transport proteins (p. 473) and has one principal fate: *it is phosphorylated by ATP to form glucose 6-phosphate.* This step is notable for several reasons. Glucose 6-phosphate cannot pass through the membrane because of the negative charges on the phosphoryl groups, and it is not a substrate for glucose transporters. Also, the addition of the phosphoryl group facilitates the eventual metabolism of glucose into three-carbon molecules with high-phosphoryl-transfer-potential. The transfer of the phosphoryl group from ATP to the hydroxyl group on carbon 6 of glucose is catalyzed by *hexokinase.*



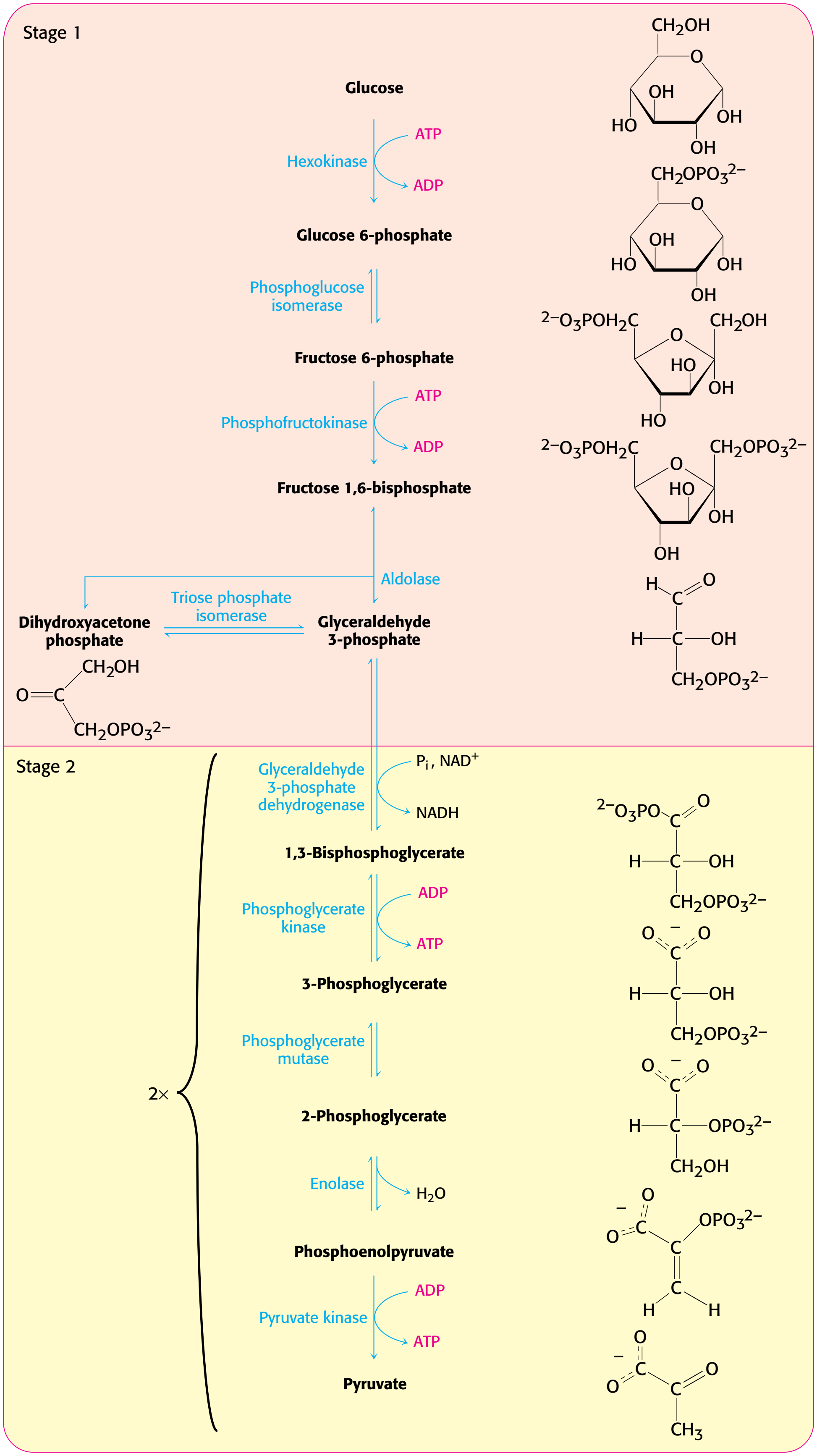

Stage 1
Glucose
Hexokinase
ATP
ADP
Glucose 6-phosphate
Phosphoglucose isomerase
Fructose 6-phosphate
Phosphofructokinase
ATP
ADP
Fructose 1,6-bisphosphate
Aldolase
Triose phosphate isomerase
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
Stage 2
2×
Glyceraldehyde 3-phosphate dehydrogenase
Pᵢ, NAD⁺
NADH
1,3-Bisphosphoglycerate
Phosphoglycerate kinase
ADP
ATP
3-Phosphoglycerate
Phosphoglycerate mutase
2-Phosphoglycerate
Enolase
H₂O
Phosphoenolpyruvate
Pyruvate kinase
ADP
ATP
Pyruvate
CH₂OH
CH₂OPO₃²⁻
²⁻O₃POH₂C
HO
OH
O
H
C
H—C—OH
²⁻O₃PO
OPO₃²⁻
CH₃

FIGURE 16.2 **Stages of glycolysis.** The glycolytic pathway can be divided into two stages: (1) glucose is trapped, destabilized, and cleaved into two interconvertible three-carbon molecules generated by cleavage of six-carbon fructose; and (2) ATP is generated.

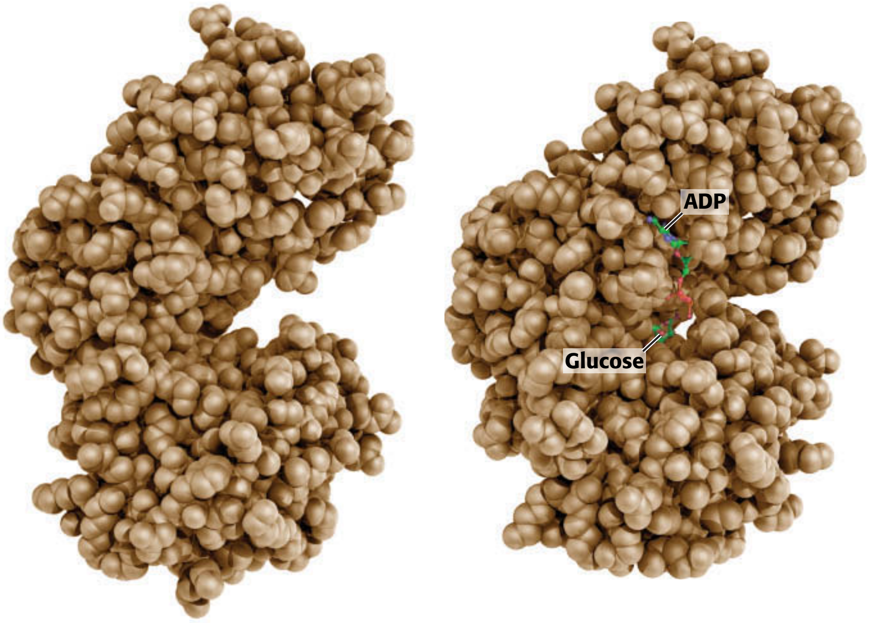


FIGURE 16.3 Induced fit in hexokinase. The two lobes of hexokinase are separated in the absence of glucose (left). The conformation of hexokinase changes markedly on binding glucose (right). *Notice* that two lobes of the enzyme come together, creating the necessary environment for catalysis. [After RSCB Protein Data Bank; drawn from yhx and 1hkg by Adam Steinberg.]

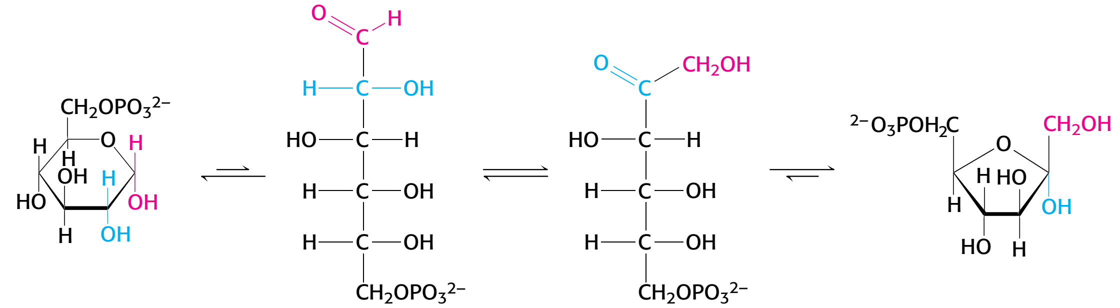
Phosphoryl transfer is a fundamental reaction in biochemistry. *Kinases* are enzymes that catalyze the transfer of a phosphoryl group from ATP to an acceptor. Hexokinase, then, catalyzes the transfer of a phosphoryl group from ATP to a variety of six-carbon sugars (*hexoses*), such as glucose and mannose. *Hexokinase, like adenylate kinase* (Section 9.4) *and all other kinases, requires* Mg^{2+} *(or another divalent metal ion such as* Mn^{2+}*) for activity.* The divalent metal ion forms a complex with ATP.

X-ray crystallographic studies of yeast hexokinase revealed that the binding of glucose induces a large conformational change in the enzyme. Hexokinase consists of two lobes, which move toward each other when glucose is bound (Figure 16.3). On glucose binding, one lobe rotates 12 degrees with respect to the other, resulting in movements of the polypeptide backbone of as much as 8 Å. The cleft between the lobes closes, and the bound glucose becomes surrounded by protein, except for the hydroxyl group of carbon 6, which will accept the phosphoryl group from ATP. The closing of the cleft in hexokinase is a striking example of the role *of induced fit* in enzyme action (Section 8.3).

The glucose-induced structural changes are significant in two respects. First, the environment around the glucose becomes more nonpolar, which favors reaction between the hydrophilic hydroxyl group of glucose and the terminal phosphoryl group of ATP. Second, the conformational changes enable the kinase to discriminate against H_2O as a substrate. The closing of the cleft keeps water molecules away from the active site. If hexokinase were rigid, a molecule of H_2O occupying the binding site for the —CH_2OH of glucose could attack the γ phosphoryl group of ATP, forming ADP and P_i. In other words, a rigid kinase would likely also be an ATPase. It is interesting to note that other kinases taking part in glycolysis—phosphofructokinase, phosphoglycerate kinase, and pyruvate kinase—also contain clefts between lobes that close when substrate is bound, although the structures of these enzymes are different in other regards. *Substrate-induced cleft closing is a general feature of kinases.* Recall that protein kinase A also undergoes similar structural changes (p. 298).

Fructose 1,6-bisphosphate is generated from glucose 6-phosphate

The next step in glycolysis is the *isomerization of glucose 6-phosphate to fructose 6-phosphate.* Recall that the open-chain form of glucose has an aldehyde group at carbon 1, whereas the open-chain form of fructose has a keto group at carbon 2. Thus, the isomerization of glucose 6-phosphate to fructose 6-phosphate is a *conversion of an aldose into a ketose.* The reaction catalyzed by *phosphoglucose isomerase* takes several steps because both glucose 6-phosphate and fructose 6-phosphate are present primarily in the cyclic forms. The enzyme must first open the six-membered ring of glucose 6-phosphate, catalyze the isomerization, and then promote the formation of the five-membered ring of fructose 6-phosphate.



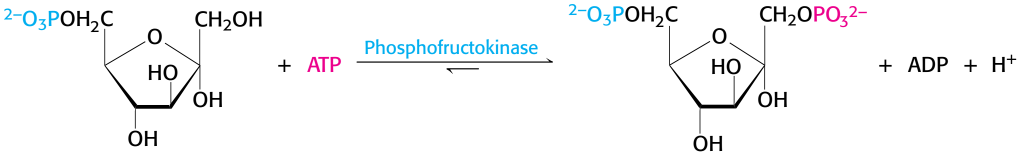
Glucose 6-phosphate (G-6P)

Glucose 6-phosphate (open-chain form)

Fructose 6-phosphate (open-chain form)

Fructose 6-phosphate (F-6P)

A second phosphorylation reaction follows the isomerization step. *Fructose 6-phosphate is phosphorylated at the expense of ATP to fructose 1,6-bisphosphate* (F-1,6-BP). The prefix *bis*- in bisphosphate means that two separate monophosphoryl groups are present, whereas the prefix *di*- in diphosphate (as in adenosine diphosphate) means that two phosphoryl groups are present and are connected by an anhydride bond.



Fructose 6-phosphate (F-6P)

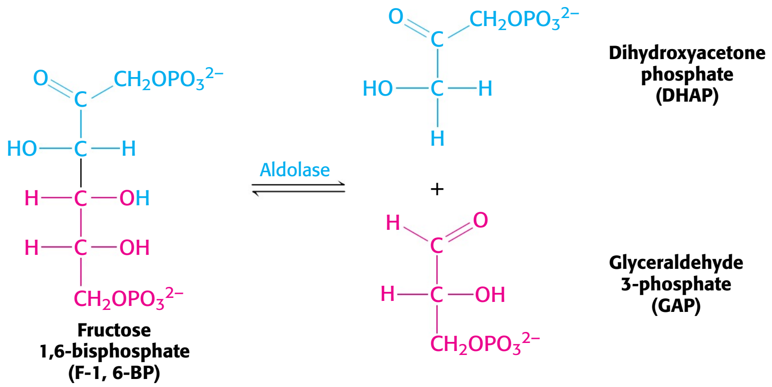
Fructose 1,6-bisphosphate (F-1, 6-BP)

This reaction is catalyzed by *phosphofructokinase* (PFK), an allosteric enzyme that sets the pace of glycolysis. As we will learn, this enzyme plays a central role in the metabolism of many molecules in all parts of the body.

What is the biochemical rationale for the isomerization of glucose 6-phosphate to fructose 6-phosphate and its subsequent phosphorylation to form fructose 1,6-bisphosphate? Had the aldol cleavage taken place in the aldose glucose, a two-carbon and a four-carbon fragment would have resulted. Two different metabolic pathways, one to process the two-carbon fragment and one for the four-carbon fragment, would have been required to extract energy. Phosphorylation of the fructose 6-phosphate to fructose 1,6-bisphosphate prevents the reformation of glucose 6-phosphate. As shown below, aldol cleavage of fructose 1,6-bisphosphate yields two phosphorylated interconvertible three-carbon fragments that will be oxidized in the later steps of glycolysis to capture energy in the form of ATP.

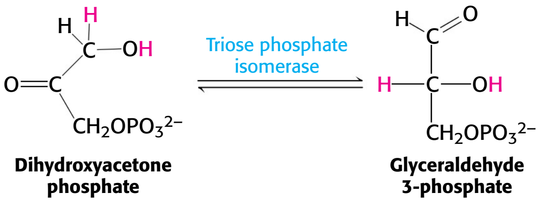
The six-carbon sugar is cleaved into two three-carbon fragments

The newly formed fructose 1,6-bisphosphate is cleaved into *glyceraldehyde 3-phosphate* (GAP) and *dihydroxyacetone phosphate* (DHAP), completing stage 1 of glycolysis. The products of the remaining steps in glycolysis consist of three-carbon units rather than six-carbon units.


Fructose 1,6-bisphosphate (F-1, 6-BP)
Aldolase
Dihydroxyacetone phosphate (DHAP)
+
Glyceraldehyde 3-phosphate (GAP)

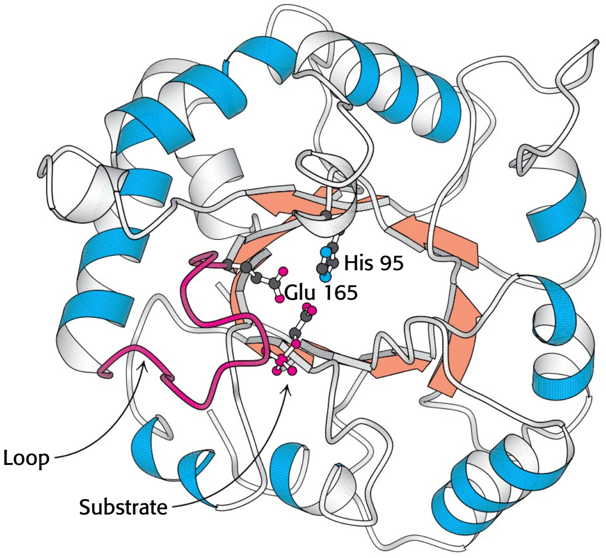
This reaction, which is readily reversible, is catalyzed by *aldolase.* This enzyme derives its name from the nature of the reverse reaction, an aldol condensation.

Glyceraldehyde 3-phosphate is on the direct pathway of glycolysis, whereas dihydroxyacetone phosphate is not. Unless a means exists to convert dihydroxyacetone phosphate into glyceraldehyde 3-phosphate, a three-carbon fragment useful for generating ATP will be lost. These compounds are isomers that can be readily interconverted: dihydroxyacetone phosphate is a ketose, whereas glyceraldehyde 3-phosphate is an aldose. The isomerization of these three-carbon phosphorylated sugars is catalyzed by *triose phosphate isomerase* (TPI, sometimes abbreviated TIM; Figure 16.4).


Dihydroxyacetone phosphate
Triose phosphate isomerase
Glyceraldehyde 3-phosphate

This reaction is rapid and reversible. At equilibrium, 96% of the triose phosphate is dihydroxyacetone phosphate. However, the reaction proceeds readily from dihydroxyacetone phosphate to glyceraldehyde 3-phosphate because the subsequent reactions of glycolysis remove this product. Triose phosphate isomerase deficiency, a rare condition, is the only glycolytic enzymopathy that is lethal. This deficiency is characterized by severe hemolytic anemia and neurodegeneration.

FIGURE 16.4 Structure of triose phosphate isomerase. This enzyme consists of a central core of eight parallel β strands (orange) surrounded by eight α helices (blue). This structural motif, called an αβ barrel, is also found in the glycolytic enzymes aldolase, enolase, and pyruvate kinase. *Notice* that histidine 95 and glutamate 165, essential components of the active site of triose phosphate isomerase, are located in the barrel. A loop (red) closes off the active site on substrate binding. [Drawn from 2YPI.pdb.]



Mechanism: Triose phosphate isomerase salvages a three-carbon fragment

Much is known about the catalytic mechanism of triose phosphate isomerase. TPI catalyzes the transfer of a hydrogen atom from carbon 1 to carbon 2, an intramolecular oxidation–reduction. This isomerization of a ketose into an aldose proceeds through an *enediol intermediate* (Figure 16.5).

X-ray crystallographic and other studies showed that glutamate 165 plays the role of a general acid–base catalyst: it abstracts a proton (H^+) from carbon 1 and then donates it to carbon 2. However, the carboxylate group of glutamate 165 by itself is not basic enough to pull a proton away from a carbon atom adjacent to a carbonyl group. Histidine 95 assists catalysis by donating a proton to stabilize the negative charge that develops on the C-2 carbonyl group.

Two features of this enzyme are noteworthy. First, TPI displays great catalytic prowess. It accelerates isomerization by a

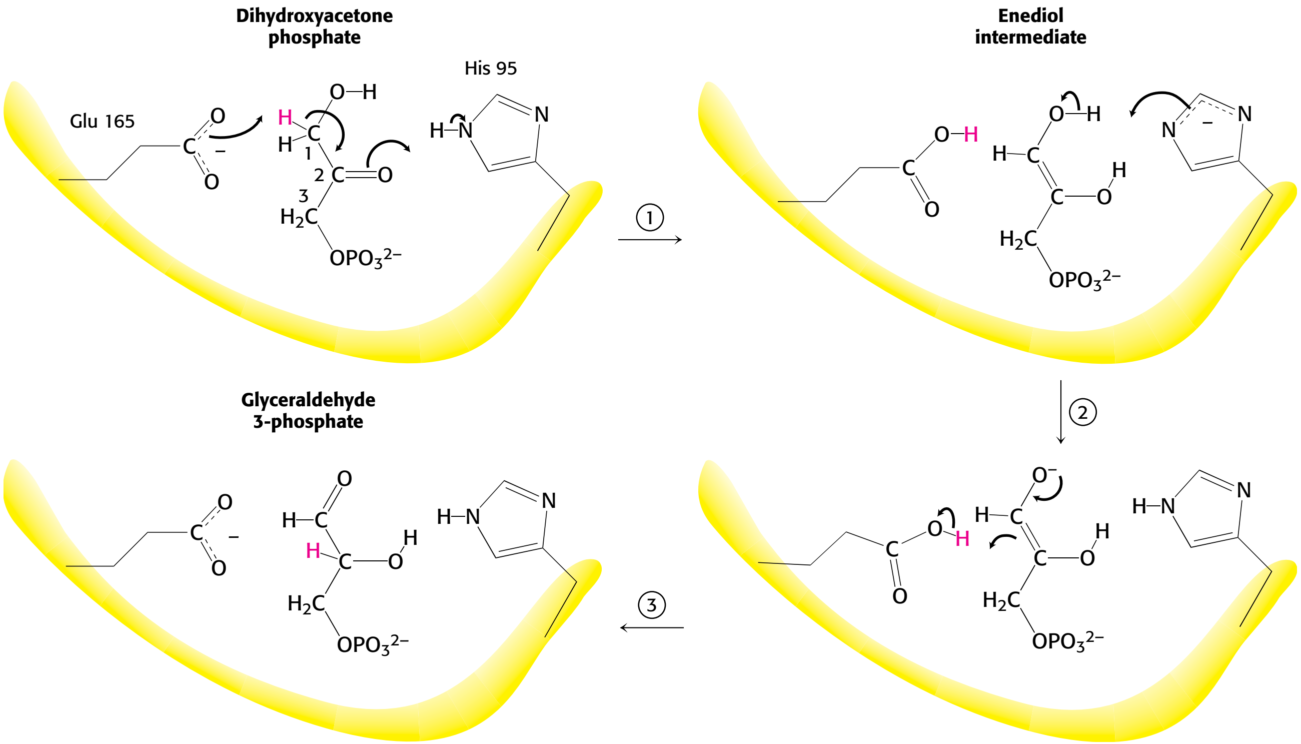
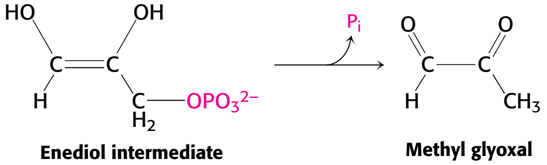


FIGURE 16.5 Catalytic mechanism of triose phosphate isomerase. (1) Glutamate 165 acts as a general base by abstracting a proton (H^+) from carbon 1. Histidine 95, acting as a general acid, donates a proton to the oxygen atom bonded to carbon 2, forming the enediol intermediate. (2) Glutamic acid, now acting as a general acid, donates a proton to C-2 while histidine removes a proton from the OH group of C-1. (3) The product is formed, and glutamate and histidine are returned to their ionized and neutral forms, respectively.

factor of 10^{10} compared with the rate obtained with a simple base catalyst such as acetate ion. Indeed, the k_{cat}/K_M ratio for the isomerization of glyceraldehyde 3-phosphate is $2 \times 10^8\ M^{-1}\ s^{-1}$, which is close to the diffusion-controlled limit. In other words, catalysis takes place every time that enzyme and substrate meet. The diffusion-controlled encounter of substrate and enzyme is thus the rate-limiting step in catalysis. TPI is an example of a *kinetically perfect enzyme* (Section 8.4). Second, TPI suppresses an undesired side reaction, the decomposition of the enediol intermediate into methyl glyoxal and orthophosphate.

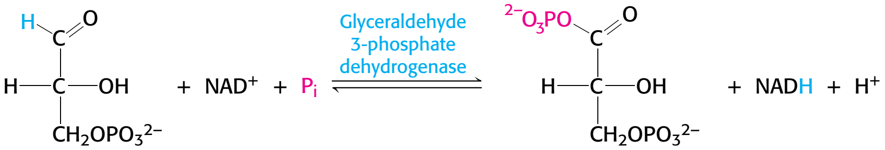


In solution, this physiologically useless reaction is 100 times as fast as isomerization. Moreover, methyl glyoxal is a highly reactive compound that can modify the structure and function of a variety of biomolecules, including proteins and DNA. The reaction of methyl glyoxal with a biomolecule is an example of deleterious reactions called advanced glycation end products, discussed earlier (AGEs, Section 11.1). Hence, TPI must prevent the enediol from leaving the enzyme. This labile intermediate is trapped in the active site by the movement of a loop of 10 residues (Figure 16.4). This loop serves as a lid on the active site, shutting it when the enediol is present and reopening it when isomerization is completed. *We see here a striking example of one means of preventing an undesirable alternative reaction: the active site is kept closed until the desirable reaction takes place.*

Thus, two molecules of glyceraldehyde 3-phosphate are formed from one molecule of fructose 1,6-bisphosphate by the sequential action of aldolase and triose phosphate isomerase. The economy of metabolism is evident in this reaction sequence. The isomerase funnels dihydroxyacetone phosphate into the main glycolytic pathway; a separate set of reactions is not needed.

The oxidation of an aldehyde to an acid powers the formation of a compound with high phosphoryl-transfer potential

The preceding steps in glycolysis have transformed one molecule of glucose into two molecules of glyceraldehyde 3-phosphate, but no energy has yet been extracted. On the contrary, thus far, two molecules of ATP have been invested. We come now to the second stage of glycolysis, a series of steps that harvest some of the energy contained in glyceraldehyde 3-phosphate as ATP. The initial reaction in this sequence is the *conversion of glyceraldehyde 3-phosphate into 1,3-bisphosphoglycerate* (1,3-BPG), a reaction catalyzed by *glyceraldehyde 3-phosphate dehydrogenase.*

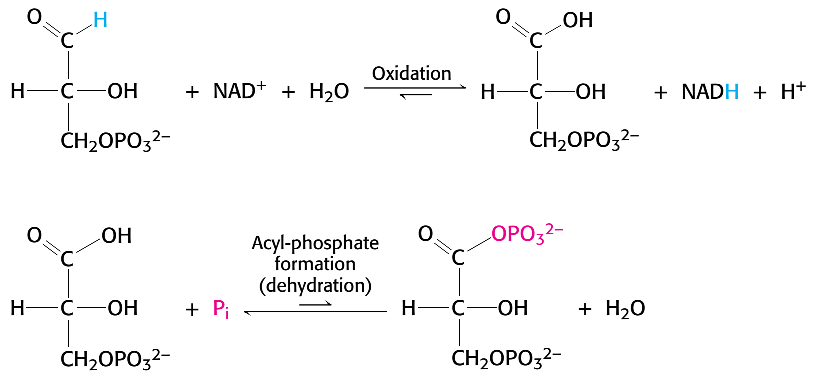

H
O
C
H—C—OH
CH2OPO3 2−
+ NAD+ + Pi
Glyceraldehyde
3-phosphate
dehydrogenase
2−O3PO
O
C
H—C—OH
CH2OPO3 2−
+ NADH + H+

Glyceraldehyde 3-phosphate (GAP)

1,3-Bisphosphoglycerate (1,3-BPG)

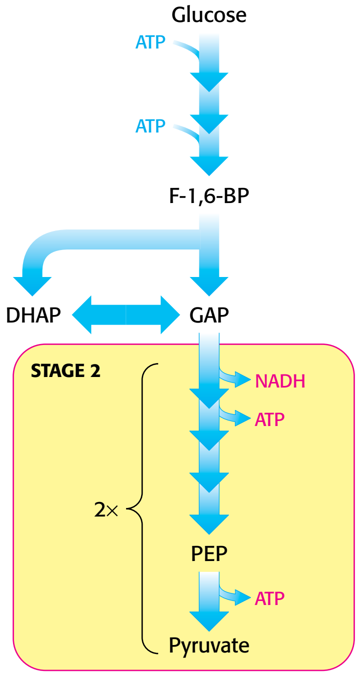
1,3-Bisphosphoglycerate is an acyl phosphate, which is a mixed anhydride of phosphoric acid and a carboxylic acid. Such compounds have a high phosphoryl-transfer potential; one of its phosphoryl groups is transferred to ADP in the next step in glycolysis.

The reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase can be viewed as the sum of two processes: the *oxidation* of the aldehyde to a carboxylic acid by NAD$^+$ and the *joining* of the carboxylic acid and orthophosphate to form the acyl-phosphate product.


O
H
C
H—C—OH
CH2OPO3 2−
+ NAD+ + H2O
Oxidation
O
OH
C
H—C—OH
CH2OPO3 2−
+ NADH + H+
O
OH
C
H—C—OH
CH2OPO3 2−
+ Pi
Acyl-phosphate
formation
(dehydration)
O
OPO3 2−
C
H—C—OH
CH2OPO3 2−
+ H2O

The first reaction is thermodynamically quite favorable, with a standard free-energy change, $\Delta G^{\circ\prime}$, of approximately -50 kJ mol^{-1} (-12 kcal mol^{-1}), whereas the second reaction is quite unfavorable, with a standard free-energy change of the same magnitude but the opposite sign. If these two reactions simply took place in succession, the second reaction would have a very large activation energy and thus not take place at a biologically significant rate. These two processes *must be coupled* so that the favorable aldehyde oxidation can be used to drive the formation of the acyl phosphate. How are these reactions coupled? *The key is an intermediate, formed as a result of the aldehyde oxidation, that is linked to the enzyme by a thioester bond.* Thioesters are high-energy compounds found in many biochemical pathways (Section 15.4). This intermediate reacts with orthophosphate to form the high-energy compound 1,3-bisphosphoglycerate.

The thioester intermediate is higher in free energy than the free carboxylic acid is. The favorable oxidation and unfavorable phosphorylation reactions are coupled by the thioester intermediate, which preserves much of the free


Glucose
ATP
ATP
F-1,6-BP
DHAP
GAP
STAGE 2
NADH
ATP
2×
PEP
ATP
Pyruvate

Second stage of glycolysis. The oxidation of three-carbon fragments yields ATP.

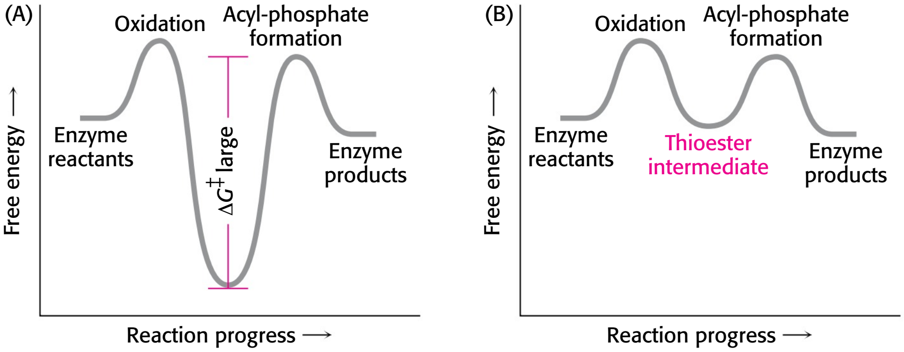


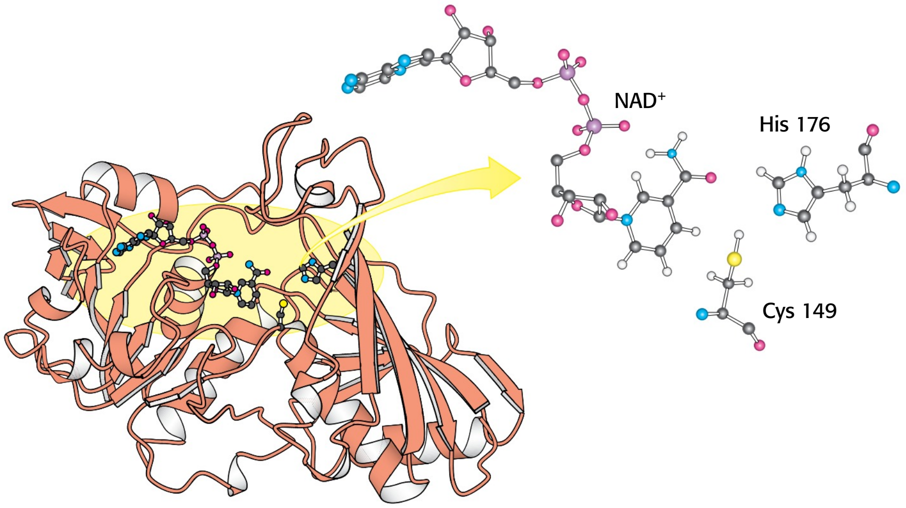
FIGURE 16.6 Free-energy profiles for glyceraldehyde oxidation followed by acyl-phosphate formation. (A) A hypothetical case with no coupling between the two processes. The second step must have a large activation barrier, making the reaction very slow. (B) The actual case with the two reactions coupled through a thioester intermediate.

energy released in the oxidation reaction. We see here the *use of a covalent enzyme-bound intermediate as a mechanism of energy coupling.* A free-energy profile of the glyceraldehyde 3-phosphate dehydrogenase reaction, compared with a hypothetical process in which the reaction proceeds without this intermediate, reveals how this intermediate allows a favorable process to drive an unfavorable one (Figure 16.6).

Mechanism: Phosphorylation is coupled to the oxidation of glyceraldehyde 3-phosphate by a thioester intermediate

The active site of glyceraldehyde 3-phosphate dehydrogenase includes a reactive cysteine residue, as well as NAD^+ and a crucial histidine (Figure 16.7). Let us consider in detail how these components cooperate in the reaction mechanism (Figure 16.8). In step 1, the aldehyde substrate reacts with the sulfhydryl group of cysteine 149 on the enzyme to form a hemithioacetal. Step 2 is the *transfer of a hydride ion to a molecule of NAD^+ that is tightly bound to the enzyme and is adjacent to the cysteine residue.* This reaction is favored by the deprotonation of the hemithioacetal by histidine 176. The products of this reaction are the reduced coenzyme NADH and a thioester intermediate. *This thioester intermediate has a free energy close to that of the reactants* (Figure 16.6). In step 3, the NADH formed from the aldehyde oxidation leaves the enzyme and is replaced by a second molecule of NAD^+. This step is important because the positive charge on NAD^+ polarizes the thioester intermediate to facilitate the

FIGURE 16.7 Structure of glyceraldehyde 3-phosphate dehydrogenase. *Notice* that the active site includes a cysteine residue and a histidine residue adjacent to a bound NAD^+ molecule. The sulfur atom of cysteine will link with the substrate to form a transitory thioester intermediate. [Drawn from 1GAD.pdb.]



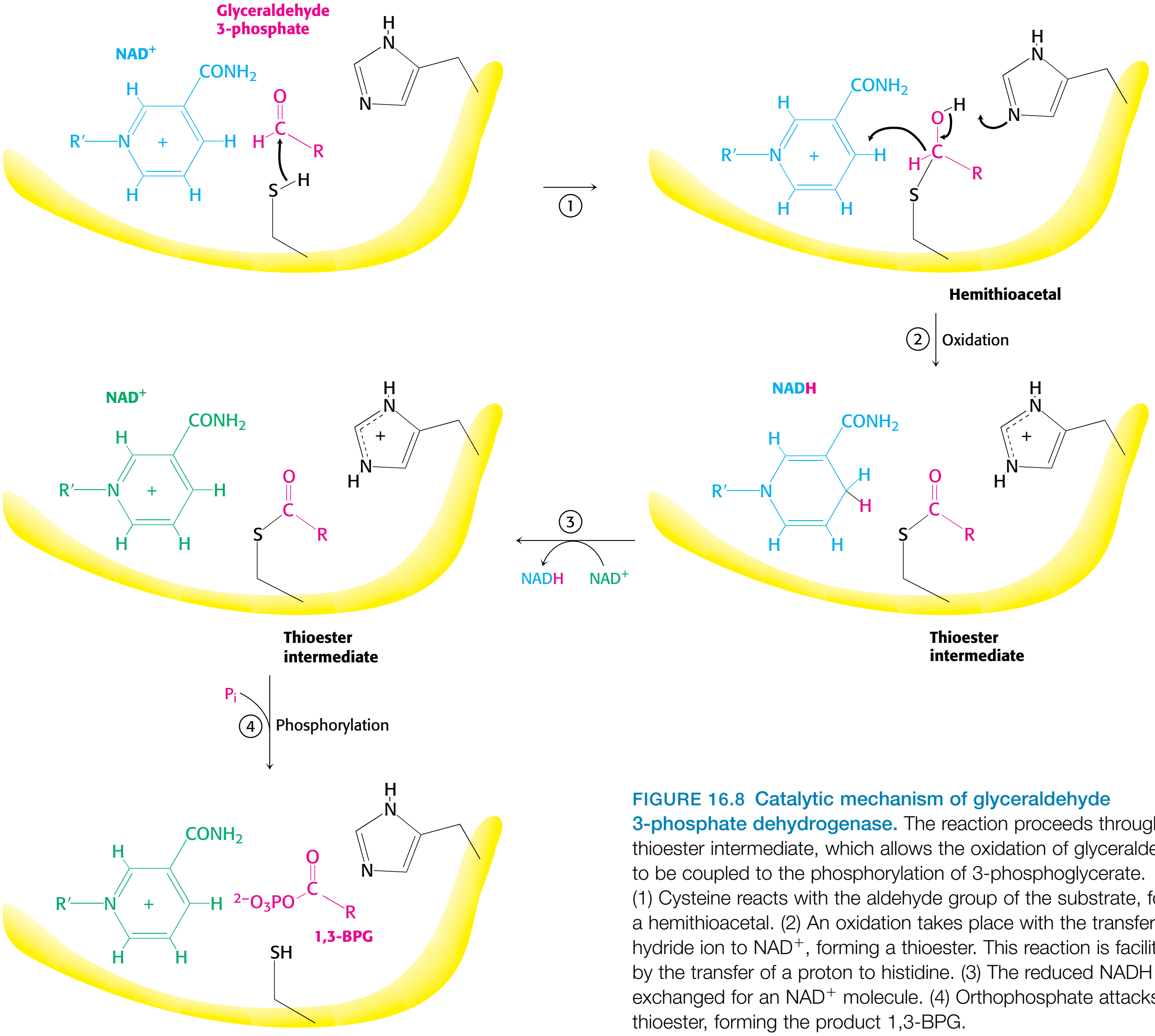
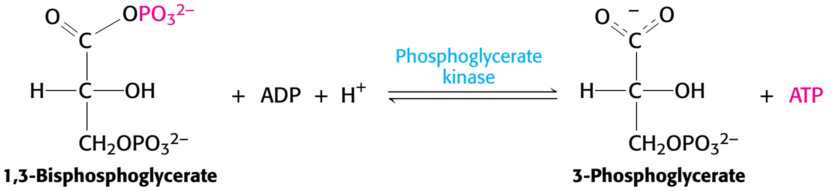

Glyceraldehyde 3-phosphate
NAD+
CONH2
R′
N
H
S
O
C
R
①
Hemithioacetal
② Oxidation
NADH
Thioester intermediate
③
NADH
NAD+
Thioester intermediate
Pi
④ Phosphorylation
2−O3PO
1,3-BPG
SH

FIGURE 16.8 Catalytic mechanism of glyceraldehyde 3-phosphate dehydrogenase. The reaction proceeds through a thioester intermediate, which allows the oxidation of glyceraldehyde to be coupled to the phosphorylation of 3-phosphoglycerate. (1) Cysteine reacts with the aldehyde group of the substrate, forming a hemithioacetal. (2) An oxidation takes place with the transfer of a hydride ion to NAD$^+$, forming a thioester. This reaction is facilitated by the transfer of a proton to histidine. (3) The reduced NADH is exchanged for an NAD$^+$ molecule. (4) Orthophosphate attacks the thioester, forming the product 1,3-BPG.

attack by orthophosphate. In step 4, orthophosphate attacks the thioester to form 1,3-BPG and free the cysteine residue. This example illustrates the essence of energy transformations and of metabolism itself: energy released by carbon oxidation is converted into high phosphoryl-transfer potential.

ATP is formed by phosphoryl transfer from 1,3-bisphosphoglycerate

1,3-Bisphosphoglycerate is an energy-rich molecule with a greater phosphoryl-transfer potential than that of ATP (Section 15.2). Thus, 1,3-BPG can be used to power the synthesis of ATP from ADP. *Phosphoglycerate kinase* catalyzes the transfer of the phosphoryl group from the acyl phosphate of 1,3-bisphosphoglycerate to ADP. ATP and 3-phosphoglycerate are the products.


O
OPO3 2−
C
H—C—OH
CH2OPO3 2−
1,3-Bisphosphoglycerate
+ ADP + H+
Phosphoglycerate kinase
O − O
C
H—C—OH
CH2OPO3 2−
3-Phosphoglycerate
+ ATP

The formation of ATP in this manner is referred to as *substrate-level phosphorylation* because the phosphate donor, 1,3-BPG, is a substrate with high phosphoryl-transfer potential. We will contrast this manner of ATP formation with the formation of ATP from ionic gradients in chapters 18 and 19.

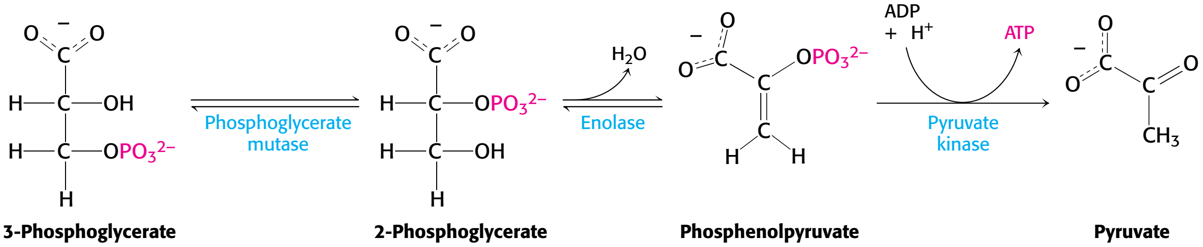
Thus, the outcomes of the reactions catalyzed by glyceraldehyde 3-phosphate dehydrogenase and phosphoglycerate kinase are as follows:

1. Glyceraldehyde 3-phosphate, an aldehyde, is oxidized to 3-phosphoglycerate, a carboxylic acid.
2. NAD^+ is concomitantly reduced to NADH.
3. ATP is formed from P_i and ADP at the expense of carbon-oxidation energy.

In essence, the energy released during the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate is temporarily trapped as 1,3-bisphosphoglycerate. This energy powers the transfer of a phosphoryl group from 1,3-bisphosphoglycerate to ADP to yield ATP. Keep in mind that, because of the actions of aldolase and triose phosphate isomerase, two molecules of glyceraldehyde 3-phosphate were formed and hence two molecules of ATP were generated. These ATP molecules make up for the two molecules of ATP consumed in the first stage of glycolysis.

Additional ATP is generated with the formation of pyruvate

In the remaining steps of glycolysis, 3-phosphoglycerate is converted into pyruvate, and a second molecule of ATP is formed from ADP.



The first reaction is a rearrangement. The position of the phosphoryl group shifts in the *conversion of 3-phosphoglycerate into 2-phosphoglycerate,* a reaction catalyzed by *phosphoglycerate mutase.* In general, a *mutase* is an enzyme that catalyzes the intramolecular shift of a chemical group, such as a phosphoryl group. The phosphoglycerate mutase reaction has an interesting mechanism: the phosphoryl group is not simply moved from one carbon atom to another. This enzyme requires catalytic amounts of 2,3-bisphosphoglycerate (2,3-BPG) to maintain an active-site histidine residue in a phosphorylated form. This phosphoryl group is transferred to 3-phosphoglycerate to reform 2,3-bisphosphoglycerate.

$$\text{Enz-His-phosphate} + \text{3-phosphoglycerate} \rightleftharpoons \text{Enz-His} + \text{2,3-bisphosphoglycerate}$$

The mutase then functions as a phosphatase: it converts 2,3-bisphosphoglycerate into 2-phosphoglycerate. The mutase retains the phosphoryl group to regenerate the modified histidine.

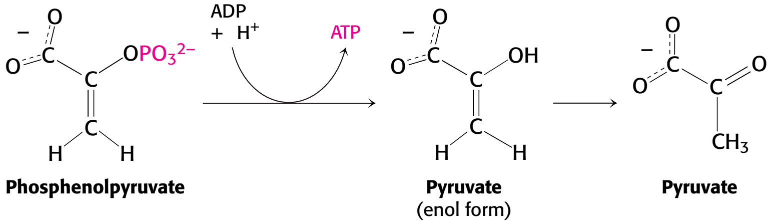
$$\text{Enz-His} + \text{2,3-bisphosphoglycerate} \rightleftharpoons \text{Enz-His-phosphate} + \text{2-phosphoglycerate}$$

The sum of these reactions yields the mutase reaction:

$$\text{3-Phosphoglycerate} \rightleftharpoons \text{2-phosphoglycerate}$$

In the next reaction, the dehydration of 2-phosphoglycerate introduces a double bond, creating an *enol. Enolase* catalyzes this formation of the enol phosphate *phosphoenolpyruvate* (PEP). This dehydration markedly elevates the transfer potential of the phosphoryl group. An *enol phosphate* has a high phosphoryl-transfer potential, whereas the phosphate ester of an ordinary alcohol, such as 2-phosphoglycerate, has a low one. The $\Delta G^{\circ\prime}$ of the hydrolysis of a phosphate ester of an ordinary alcohol is -13 kJ mol^{-1} (-3 kcal mol^{-1}), whereas that of phosphoenolpyruvate is -62 kJ mol^{-1} (-15 kcal mol^{-1}).

Why does phosphoenolpyruvate have such a high phosphoryl-transfer potential? The phosphoryl group traps the molecule in its unstable enol form. When the phosphoryl group has been donated to ATP, the enol undergoes a conversion into the more stable ketone—namely, pyruvate.



Thus, *the high phosphoryl-transfer potential of phosphoenolpyruvate arises primarily from the large driving force of the subsequent enol–ketone conversion.* Hence, pyruvate is formed, and ATP is generated concomitantly. The virtually irreversible transfer of a phosphoryl group from phosphoenolpyruvate to ADP is catalyzed *by pyruvate kinase.* What is the energy source for the formation of phosphoenolpyruvate? The answer to this question becomes clear when we compare the structures of 2-phosphoglycerate and pyruvate. The formation of pyruvate from 2-phosphoglycerate is, in essence, an internal oxidation–reduction reaction; carbon 3 takes electrons from carbon 2 in the conversion of 2-phosphoglycerate into pyruvate. Compared with 2-phosphoglycerate, C-3 is more reduced in pyruvate, whereas C-2 is more oxidized. Once again, carbon oxidation powers the synthesis of a compound with high phosphoryl-transfer potential, phosphoenolpyruvate here and 1,3-bisphosphoglycerate earlier, which allows the synthesis of ATP.

Because the molecules of ATP used in forming fructose 1,6-bisphosphate have already been regenerated, the two molecules of ATP generated from phosphoenolpyruvate are "profit."

Two ATP molecules are formed in the conversion of glucose into pyruvate

The net reaction in the transformation of glucose into pyruvate is

$$\text{Glucose} + 2\ \text{P}_i + 2\ \text{ADP} + 2\ \text{NAD}^+ \rightarrow 2\ \text{pyruvate} + 2\ \text{ATP} + 2\ \text{NADH} + 2\ \text{H}^+ + 2\ \text{H}_2\text{O}$$

Thus, *two molecules of ATP are generated in the conversion of glucose into two molecules of pyruvate.* The reactions of glycolysis are summarized in Table 16.1.

The energy released in the anaerobic conversion of glucose into two molecules of pyruvate is about -90 kJ mol^{-1} (-22 kcal mol^{-1}). We shall

TABLE 16.1 Reactions of glycolysis

Step	Reaction
1	Glucose + ATP $\longrightarrow$ glucose 6-phosphate + ADP + H^+
2	Glucose 6-phosphate $\rightleftharpoons$ fructose 6-phosphate
3	Fructose 6-phosphate + ATP $\longrightarrow$ fructose 1,6-bisphosphate + ADP + H^+
4	Fructose 1,6-bisphosphate $\rightleftharpoons$ dihydroxyacetone phosphate + glyceraldehyde 3-phosphate
5	Dihydroxyacetone phosphate $\rightleftharpoons$ glyceraldehyde 3-phosphate
6	Glyceraldehyde 3-phosphate + P_i + NAD^+ $\rightleftharpoons$ 1,3-bisphosphoglycerate + NADH + H^+
7	1,3-Bisphosphoglycerate + ADP $\rightleftharpoons$ 3-phosphoglycerate + ATP
8	3-Phosphoglycerate $\rightleftharpoons$ 2-phosphoglycerate
9	2-Phosphoglycerate $\rightleftharpoons$ phosphoenolpyruvate + H_2O
10	Phosphoenolpyruvate + ADP + H^+ $\longrightarrow$ pyruvate + ATP

Note: ΔG, the actual free-energy change, has been calculated from $\Delta G^{\circ\prime}$ and known concentrations of reactants under typical physiological conditions. Glycolysis can proceed only if the ΔG values of all reactions are negative. The small positive ΔG values of three of the above reactions indicate that the concentrations of metabolites in vivo in cells undergoing glycolysis are not precisely known.

see in chapters 17 and 18 that much more energy can be released from glucose in the presence of oxygen.

NAD^+ is regenerated from the metabolism of pyruvate

The conversion of glucose into two molecules of pyruvate has resulted in the net synthesis of ATP. However, an energy-converting pathway that stops at pyruvate will not proceed for long, because redox balance has not been maintained. As we have seen, the activity of glyceraldehyde 3-phosphate dehydrogenase, in addition to generating a compound with high phosphoryl-transfer potential, reduces NAD^+ to NADH. In the cell, there are limited amounts of NAD^+, which is derived from the vitamin niacin (B_3), a dietary requirement for human beings. Consequently, NAD^+ must be regenerated for glycolysis to proceed. Thus, the final process in the pathway is the regeneration of NAD^+ through the metabolism of pyruvate.

The sequence of reactions from glucose to pyruvate is similar in most organisms and most types of cells. In contrast, the fate of pyruvate is variable. Three reactions of pyruvate are of primary importance: conversion into ethanol, lactate, or carbon dioxide (Figure 16.9). The first two reactions are fermentations that take place in the absence of oxygen. A *fermentation* is an ATP-generating process in which organic compounds act both as the donor and as the acceptor of electrons. In the presence of oxygen, the most common situation in multicellular organisms and in many unicellular ones, pyruvate is metabolized to carbon dioxide and water through the citric acid cycle and the electron-transport chain with oxygen serving as the final electron acceptor. We now take a closer look at these three possible fates of pyruvate.

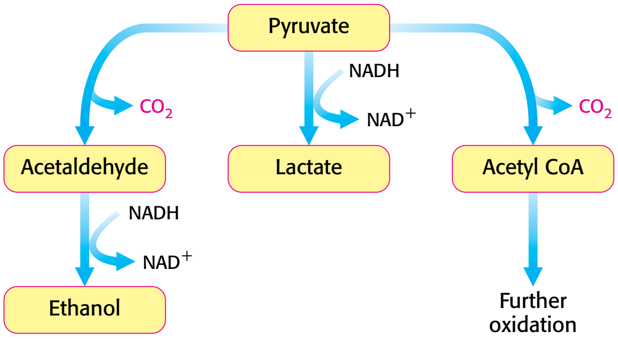
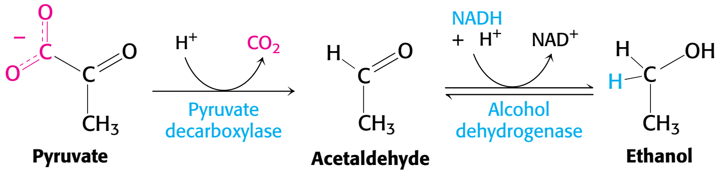


FIGURE 16.9 Diverse fates of pyruvate. Ethanol and lactate can be formed by reactions that include NADH. Alternatively, a two-carbon unit from pyruvate can be coupled to coenzyme A (Chapter 17) to form acetyl CoA.

1. *Ethanol* is formed from pyruvate in yeast and several other microorganisms. The first step is the decarboxylation of pyruvate. This reaction is catalyzed by *pyruvate decarboxylase,* which requires the coenzyme thiamine pyrophosphate, a coenzyme, derived from the vitamin thiamine (B_1). The second step is the reduction of acetaldehyde to ethanol

Enzyme	Reaction type	$\Delta G^{\circ\prime}$ in kJ mol^{-1} ($kcal\ mol^{-1}$)	ΔG in kJ mol^{-1} ($kcal\ mol^{-1}$)
Hexokinase	Phosphoryl transfer	−16.7 (−4.0)	−33.5 (−8.0)
Phosphoglucose isomerase	Isomerization	+1.7 (+0.4)	−2.5 (−0.6)
Phosphofructokinase	Phosphoryl transfer	−14.2 (−3.4)	−22.2 (−5.3)
Aldolase	Aldol cleavage	+23.8 (+5.7)	−1.3 (−0.3)
Triose phosphate isomerase	Isomerization	+7.5 (+1.8)	+2.5 (+0.6)
Glyceraldehyde 3-phosphate dehydrogenase	Phosphorylation coupled to oxidation	+6.3 (+1.5)	−1.7 (−0.4)
Phosphoglycerate kinase	Phosphoryl transfer	−18.8 (−4.5)	+1.3 (+0.3)
Phosphoglycerate mutase	Phosphoryl shift	+4.6 (+1.1)	+0.8 (+0.2)
Enolase	Dehydration	+1.7 (+0.4)	−3.3 (−0.8)
Pyruvate kinase	Phosphoryl transfer	−31.4 (−7.5)	−16.7 (−4.0)

by NADH, in a reaction catalyzed by *alcohol dehydrogenase.* This reaction regenerates NAD^+.


Pyruvate
H+
CO2
Pyruvate decarboxylase
Acetaldehyde
NADH + H+
NAD+
Alcohol dehydrogenase
Ethanol

The active site of alcohol dehydrogenase contains a zinc ion that is coordinated to the sulfur atoms of two cysteine residues and a nitrogen atom of histidine (Figure 16.10). This zinc ion polarizes the carbonyl group of the substrate to favor the transfer of a hydride from NADH.

The conversion of glucose into ethanol is an example *of alcoholic fermentation.* The net result of this anaerobic process is

$$\text{Glucose} + 2\,P_i + 2\,ADP + 2\,H^+ \rightarrow 2\,\text{ethanol} + 2\,CO_2 + 2\,ATP + 2\,H_2O$$

Note that NAD^+ and NADH do not appear in this equation, even though they are crucial for the overall process. NADH generated by the oxidation of glyceraldehyde 3-phosphate is consumed in the reduction of acetaldehyde to ethanol. Thus, *there is no net oxidation–reduction in the conversion of glucose into ethanol* (Figure 16.11). The ethanol formed in alcoholic fermentation provides a key ingredient for brewing and winemaking.

2. *Lactate* is formed from pyruvate in a variety of microorganisms in a process called *lactic acid fermentation.* Certain types of skeletal muscles

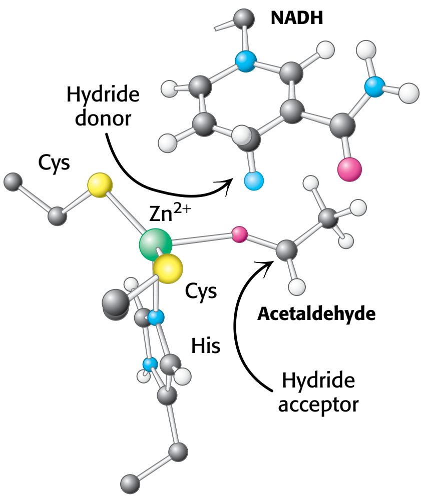
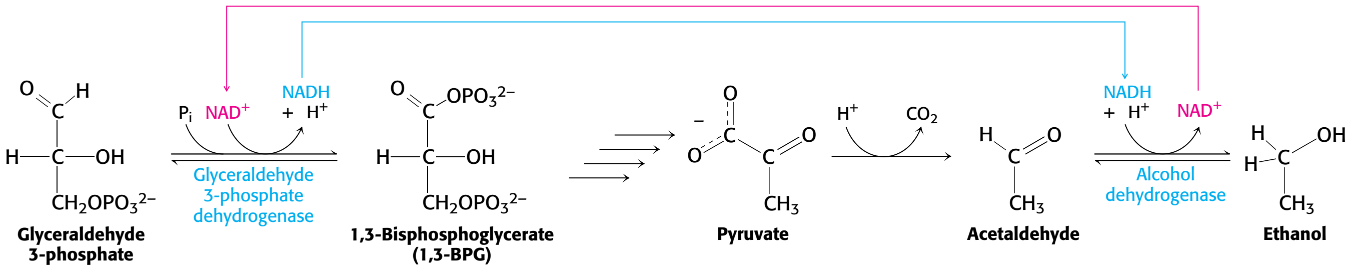
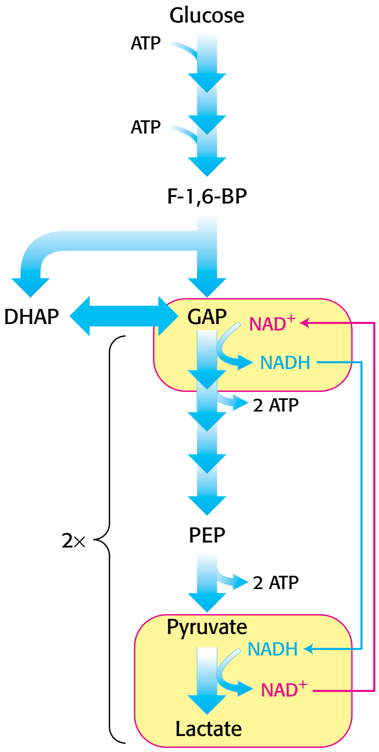

NADH
Hydride donor
Cys
Zn2+
Cys
His
Acetaldehyde
Hydride acceptor

FIGURE 16.10 Active site of alcohol dehydrogenase. The active site contains a zinc ion bound to two cysteine residues and one histidine residue. *Notice* that the zinc ion binds the acetaldehyde substrate through its oxygen atom, polarizing the substrate so that it more easily accepts a hydride from NADH. Only the nicotinamide ring of NADH is shown.

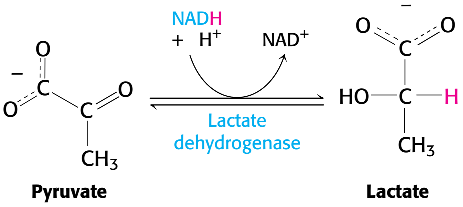
FIGURE 16.11 Maintaining redox balance. The NADH produced by the glyceraldehyde 3-phosphate dehydrogenase reaction must be reoxidized to NAD^+ for the glycolytic pathway to continue. In alcoholic fermentation, alcohol dehydrogenase oxidizes NADH and generates ethanol. In lactic acid fermentation (not shown), lactate dehydrogenase oxidizes NADH while generating lactic acid.


Glyceraldehyde 3-phosphate
Pi NAD+
NADH + H+
Glyceraldehyde 3-phosphate dehydrogenase
1,3-Bisphosphoglycerate (1,3-BPG)
Pyruvate
H+
CO2
Acetaldehyde
NADH + H+
NAD+
Alcohol dehydrogenase
Ethanol

in most animals can also function anaerobically for short periods. For example, a specific type of muscle fiber, called fast twitch or type IIb fibers, performs short bursts of intense exercise. The ATP needs rise faster than the ability of the body to provide oxygen to the muscle. The muscle functions anaerobically until fatigue sets in, which is caused, in part, by lactate buildup. Indeed, the pH of resting type IIb muscle fibers, which is about 7.0, may fall to as low as 6.3 during the bout of exercise. The drop in pH inhibits phosphofructokinase (p. 469). A lactate/H^+ symporter allows the exit of lactate from the muscle cell. The reduction of pyruvate by NADH to form lactate is catalyzed by *lactate dehydrogenase*.



Regeneration of NAD$^+$.



The overall reaction in the conversion of glucose into lactate is

$$\text{Glucose} + 2\,P_i + 2\,\text{ADP} \rightarrow 2\,\text{lactate} + 2\,\text{ATP} + 2\,H_2O$$

As in alcoholic fermentation, there is no net oxidation–reduction. The NADH formed in the oxidation of glyceraldehyde 3-phosphate is consumed in the reduction of pyruvate. *The regeneration of NAD$^+$ in the reduction of pyruvate to lactate or ethanol sustains the continued process of glycolysis under anaerobic conditions.*

3. Only a fraction of the energy of glucose is released in its anaerobic conversion into ethanol or lactate. Much more energy can be extracted aerobically by means of the citric acid cycle and the electron-transport chain. The entry point to this oxidative pathway is *acetyl coenzyme A* (acetyl CoA), which is formed inside mitochondria by the oxidative decarboxylation of pyruvate.

$$\text{Pyruvate} + \text{NAD}^+ + \text{CoA} \rightarrow \text{acetyl CoA} + CO_2 + \text{NADH} + H^+$$

This reaction, which is catalyzed by the pyruvate dehydrogenase complex, will be considered in detail in Chapter 17. The NAD$^+$ required for this reaction and for the oxidation of glyceraldehyde 3-phosphate is regenerated when NADH ultimately transfers its electrons to O_2 through the electron-transport chain in mitochondria.

Fermentations provide usable energy in the absence of oxygen

Fermentations yield only a fraction of the energy available from the complete combustion of glucose. Why is a relatively inefficient metabolic pathway so extensively used? The fundamental reason is that oxygen is not required. The ability to survive without oxygen affords a host of living accommodations such as soils, deep water, and skin pores. Some organisms, called *obligate anaerobes,* cannot survive in the presence of O_2, a highly reactive compound. The bacterium *Clostridium perfringens,* the cause of gangrene, is an example of an obligate anaerobe. Other pathogenic obligate anaerobes are listed in Table 16.2. Some organisms, such as yeast, are *facultative anaerobes* that metabolize glucose aerobically when oxygen is present and perform fermentation when oxygen is absent.

Although we have considered only lactic acid and alcoholic fermentation, microorganisms are capable of generating a wide array of molecules as

TABLE 16.2 Examples of pathogenic obligate anaerobes

Bacterium	Result of infection
Clostridium tetani	Tetanus (lockjaw)
Clostridium botulinum	Botulism (an especially severe type of food poisoning)
Clostridium perfringens	Gas gangrene (gas is produced as an end point of the fermentation, distorting and destroying the tissue)
Bartonella hensela	Cat scratch fever (flu-like symptoms)
Bacteroides fragilis	Abdominal, pelvic, pulmonary, and blood infections

TABLE 16.3 Starting and ending points of various fermentations

Glucose	→	lactate
Lactate	→	acetate
Glucose	→	ethanol
Ethanol	→	acetate
Arginine	→	carbon dioxide
Pyrimidines	→	carbon dioxide
Purines	→	formate
Ethylene glycol	→	acetate
Threonine	→	propionate
Leucine	→	2-alkylacetate
Phenylalanine	→	propionate

Note: The products of some fermentations are the substrates for others.

end points of fermentation (Table 16.3). Indeed, many food products, including sour cream, yogurt, various cheeses, beer, wine, and sauerkraut, result from fermentation.

The binding site for NAD^+ is similar in many dehydrogenases

The three dehydrogenases—glyceraldehyde 3-phosphate dehydrogenase, alcohol dehydrogenase, and lactate dehydrogenase—have quite different three-dimensional structures. However, their NAD^+-binding domains are strikingly similar (Figure 16.12). This nucleotide-binding region is made up of four α helices and a sheet of six parallel β strands. Moreover, in all cases, the bound NAD^+ displays nearly the same conformation. This common structural domain is often called a *Rossmann fold* after Michael Rossmann, who first recognized it. This fold likely represents a primordial dinucleotide-binding domain that recurs in the dehydrogenases of glycolysis and other enzymes because of their descent from a common ancestor.

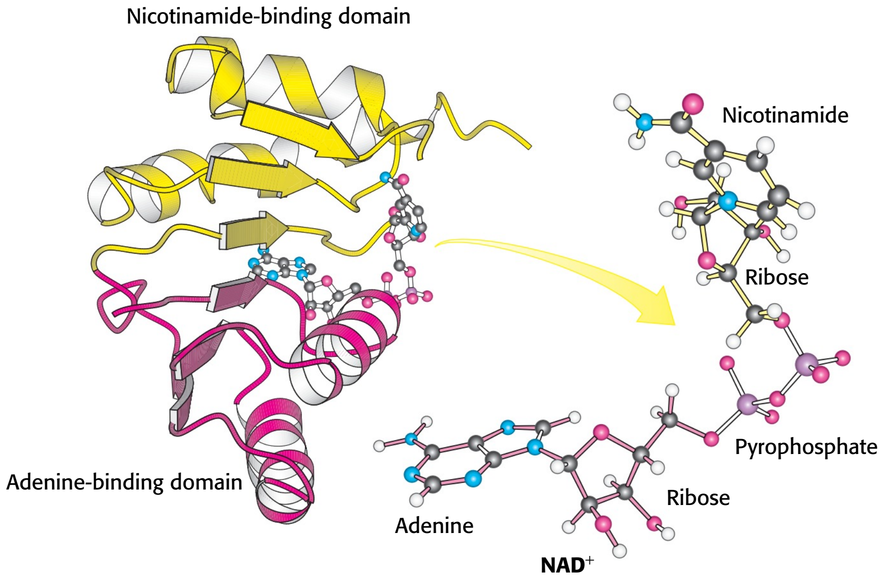


FIGURE 16.12 NAD^+-binding region in dehydrogenases. *Notice* that the nicotinamide-binding half (yellow) is structurally similar to the adenine-binding half (red). The two domains together form a structural motif called a Rossmann fold. The NAD^+ molecule binds in an extended conformation. [Drawn from 3LDH.pdb.]

Fructose is converted into glycolytic intermediates by fructokinase

Although glucose is the most widely used monosaccharide, others also are important fuels. Let us consider how fructose is funneled into the glycolytic pathway (Figure 16.13). There are no catabolic pathways for metabolizing fructose, and so the strategy is to convert this sugar into a metabolite of glucose.

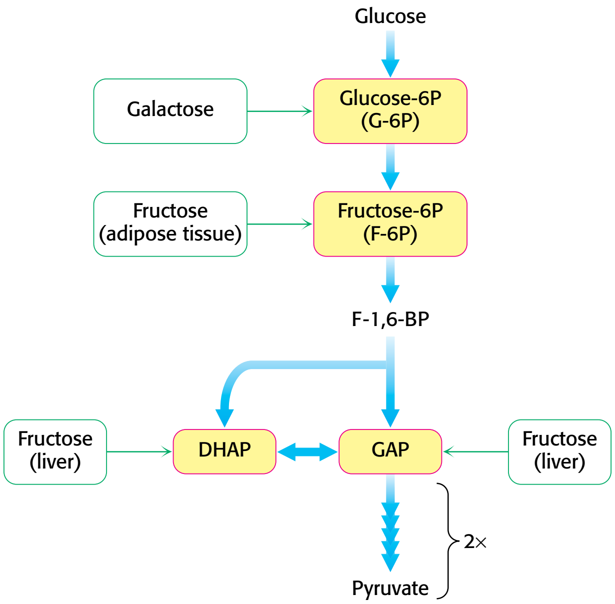

Glucose
Galactose
Glucose-6P (G-6P)
Fructose (adipose tissue)
Fructose-6P (F-6P)
F-1,6-BP
Fructose (liver)
DHAP
GAP
Fructose (liver)
2×
Pyruvate

FIGURE 16.13 Entry points in glycolysis for fructose and galactose.

The main site of fructose metabolism is the liver, using *the fructose 1-phosphate pathway* (Figure 16.14). The first step is the phosphorylation of *fructose* to *fructose 1-phosphate* by *fructokinase.* Fructose 1-phosphate is then split into *glyceraldehyde* and *dihydroxyacetone phosphate,* an intermediate in glycolysis. This aldol cleavage is catalyzed by a specific *fructose 1-phosphate aldolase.* Glyceraldehyde is then phosphorylated to *glyceraldehyde 3-phosphate,* a glycolytic intermediate, by *triose kinase.* In other tissues, such as adipose tissue, *fructose can be phosphorylated to fructose 6-phosphate by hexokinase.*

Excessive fructose consumption can lead to pathological conditions

Fructose, a commonly used sweetener, is a component of sucrose and high fructose corn syrup (which contains approximately 55% fructose and 45% glucose). Epidemiological as well as clinical studies have linked excessive fructose consumption to fatty liver, insulin insensitivity, and obesity. These conditions may eventually result in type 2 diabetes (Chapter 27). Studies have shown that these disorders are not necessarily the result of simple excess energy consumption, but rather how fructose is processed by the liver. What aspects of liver fructose metabolism are the contributing factors then? Note that, as shown in Figure 16.14, the actions of fructokinase and triose kinase bypass the most important regulatory step in glycolysis, the phosphofructokinase-catalyzed reaction. The fructose-derived glyceraldehyde 3-phosphate and dihydroxyacetone phosphate are processed by glycolysis to pyruvate and subsequently to acetyl CoA in an unregulated fashion. As we will see in Chapter 22, this excess acetyl CoA is converted to fatty acids, which can be transported to adipose tissue, resulting in obesity. The liver also begins to accumulate fatty acids, resulting in fatty liver. The activity of the fructokinase and triose kinase can deplete the liver of ATP and inorganic phosphate, compromising liver function. We will return to the topic of obesity and caloric homeostasis in Chapter 27.

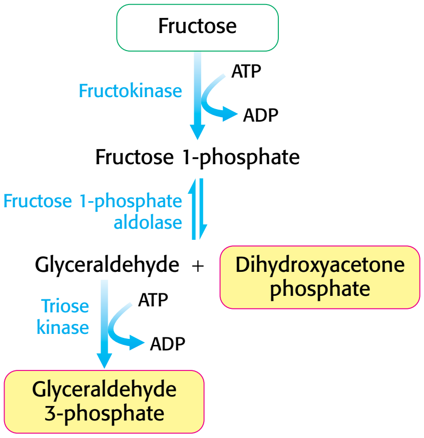
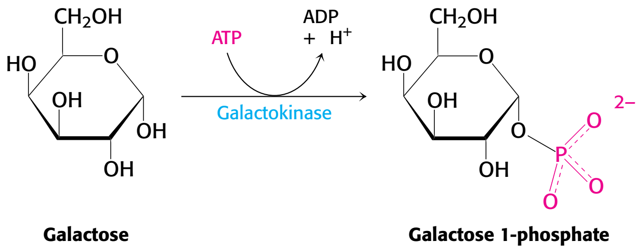

Fructose
Fructokinase
ATP
ADP
Fructose 1-phosphate
Fructose 1-phosphate aldolase
Glyceraldehyde +
Dihydroxyacetone phosphate
Triose kinase
ATP
ADP
Glyceraldehyde 3-phosphate

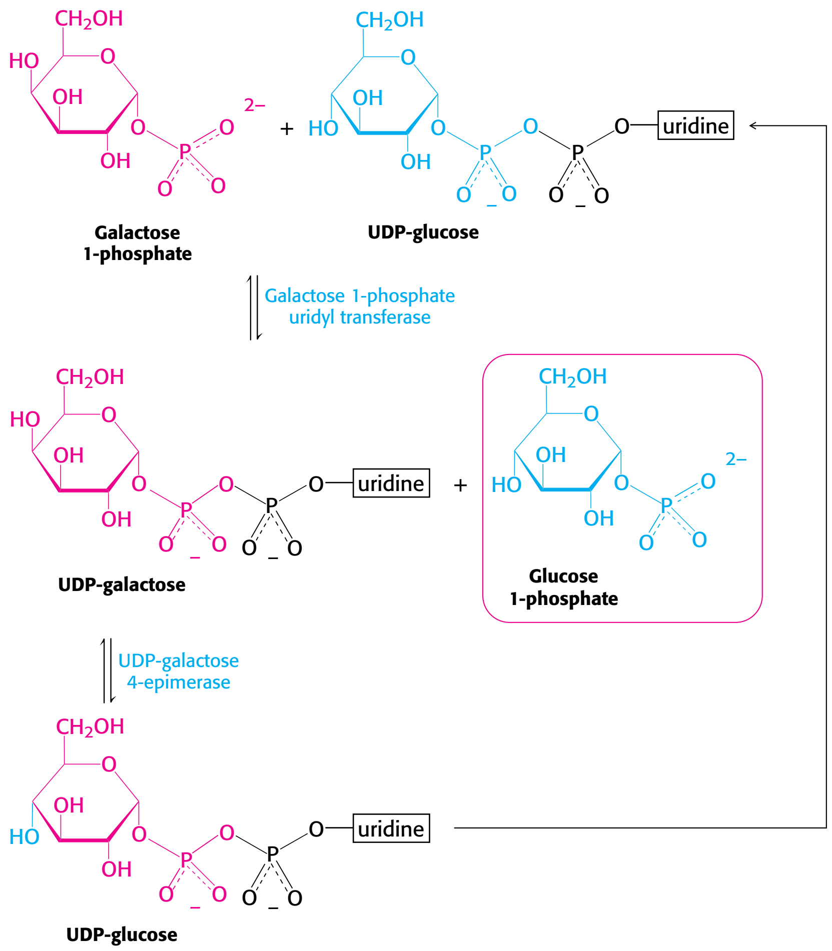
FIGURE 16.14 Fructose metabolism. Fructose enters the glycolytic pathway in the liver through the fructose 1-phosphate pathway.

Galactose is converted into glucose 6-phosphate

Like fructose, *galactose* is an abundant sugar that must be converted into metabolites of glucose (Figure 16.13). Galactose is converted into *glucose 6-phosphate* in four steps. The first reaction in the *galactose–glucose interconversion pathway* is the phosphorylation of galactose to galactose 1-phosphate by *galactokinase.*


CH2OH
HO
O
OH
OH
OH
ATP
ADP + H+
Galactokinase
CH2OH
HO
O
OH
OH
O
P
O
O
O
2–
Galactose
Galactose 1-phosphate

Galactose 1-phosphate then acquires a uridyl group from uridine diphosphate glucose (UDP-glucose), an activated intermediate in the synthesis of carbohydrates (p. 331 and Section 21.4).



The products of this reaction, which is catalyzed by *galactose 1-phosphate uridyl transferase,* are UDP-galactose and glucose 1-phosphate. The galactose moiety of UDP-galactose is then epimerized to glucose. The configuration of the hydroxyl group at carbon 4 is inverted by *UDP-galactose 4-epimerase.*

The sum of the reactions catalyzed by galactokinase, the transferase, and the epimerase is

$$\text{Galactose} + \text{ATP} \rightarrow \text{glucose 1-phosphate} + \text{ADP} + \text{H}^+$$

Note that UDP-glucose is not consumed in the conversion of galactose into glucose, because it is regenerated from UDP-galactose by the epimerase. This reaction is reversible, and the product of the reverse direction also is important. *The conversion of UDP-glucose into UDP-galactose is essential for the synthesis of galactosyl residues in complex polysaccharides and glycoproteins if the amount of galactose in the diet is inadequate to meet these needs.*

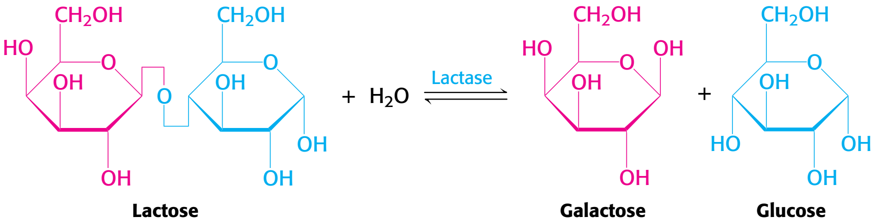
Finally, glucose 1-phosphate, formed from galactose, is isomerized to glucose 6-phosphate by *phosphoglucomutase.*

$$\text{Glucose 1-phosphate} \xrightleftharpoons{\text{Phosphoglucomutase}} \text{glucose 6-phosphate}$$

We shall return to this reaction when we consider the synthesis and degradation of glycogen, which proceeds through glucose 1-phosphate, in Chapter 21.

Many adults are intolerant of milk because they are deficient in lactase

Many adults are unable to metabolize the milk sugar lactose and experience gastrointestinal disturbances if they drink milk. *Lactose* *intolerance,* or hypolactasia, is most commonly caused by a deficiency of the enzyme lactase, which cleaves lactose into glucose and galactose.



"Deficiency" is not quite the appropriate term, because a decrease in lactase is normal in the course of development in all mammals. As children are weaned and milk becomes less prominent in their diets, lactase activity normally declines to about 5% to 10% of the level at birth. This decrease is not as pronounced with some groups of people, most notably Northern Europeans, and people from these groups can continue to ingest milk without gastrointestinal difficulties. With the development of dairy farming, an adult with active lactase would have a selective advantage in being able to consume calories from the readily available milk. Indeed, estimates suggest that people with the mutation would produce almost 20% more fertile offspring. Because dairy farming appeared in northern Europe about 10,000 years ago, the evolutionary selective pressure on lactase persistence must have been substantial, attesting to the biochemical value of being able to use milk as an energy source into adulthood.

What happens to the lactose in the intestine of a lactase-deficient person? The lactose is a good energy source for microorganisms in the colon, and they ferment it to lactic acid while generating methane (CH_4) and hydrogen gas (H_2). The gas produced creates the uncomfortable feeling of gut distension and the annoying problem of flatulence. The lactate produced by the microorganisms is osmotically active and draws water into the intestine, as does any undigested lactose, resulting in diarrhea. If severe enough, the gas and diarrhea hinder the absorption of other nutrients such as fats and proteins. The simplest treatment is to avoid the consumption of products containing much lactose. Alternatively, the enzyme lactase can be ingested with milk products.

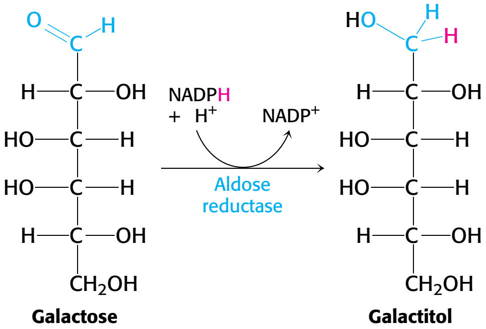
Scanning electron micrograph of *Lactobacillus*. The anaerobic bacterium *Lactobacillus* is shown here. As suggested by its name, this genus of bacteria ferments glucose into lactic acid and is widely used in the food industry. *Lactobacillus* is also a component of the normal human bacterial flora of the urogenital tract where, because of its ability to generate an acidic environment, it prevents the growth of harmful bacteria. [Power and Syred/ Science Photo Library.]

Galactose is highly toxic if the transferase is missing

Less common than lactose intolerance are disorders that interfere with the metabolism of galactose. The disruption of galactose metabolism is referred to as *galactosemia.* The most common form, called classic galactosemia, is an inherited deficiency in galactose 1-phosphate uridyl transferase activity. Afflicted infants fail to thrive. They vomit or have diarrhea after consuming milk, and enlargement of the liver and jaundice are common, sometimes progressing to cirrhosis. Cataracts will form, and lethargy and retarded mental development also are common. The blood-galactose level is markedly elevated, and galactose is found in the urine. The absence of the transferase in red blood cells is a definitive diagnostic criterion.

The most common treatment is to remove galactose (and lactose) from the diet. An enigma of galactosemia is that, although elimination of galactose from the diet prevents liver disease and cataract development, the majority of patients still suffer from central nervous system malfunction, most commonly a delayed acquisition of language skills. Female patients also display ovarian failure.

Cataract formation is better understood. A cataract is the clouding of the normally clear lens of the eye (Figure 16.15). If the transferase is not active in the lens of the eye, the presence of aldose reductase causes the accumulating galactose to be reduced to galactitol.



Galactitol is poorly metabolized and accumulates in the lens. Water will diffuse into the lens to maintain osmotic balance, triggering the formation of cataracts. In fact, there is a high incidence of cataract formation with age in populations that consume substantial amounts of milk into adulthood.

16.2 The Glycolytic Pathway Is Tightly Controlled

The glycolytic pathway has a dual role: it degrades glucose to generate ATP and it provides building blocks for biosynthetic reactions. The rate of conversion of glucose into pyruvate is regulated to meet these two major cellular needs. *In metabolic pathways, enzymes catalyzing essentially irreversible reactions are potential sites of control.* In glycolysis, the reactions catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase are virtually irreversible, and each of them serves as a control site. These enzymes become more active or less so in response to the reversible binding of allosteric effectors or to covalent modification. In addition, the amounts of these important enzymes are varied by the regulation of transcription to meet changing metabolic needs. The time required for allosteric control, regulation by phosphorylation, and transcriptional control is measured typically in milliseconds, seconds, and hours, respectively. We will consider the control of glycolysis in two different tissues—skeletal muscle and liver.

Glycolysis in muscle is regulated to meet the need for ATP

Glycolysis in skeletal muscle provides ATP primarily to power contraction. Consequently, *the primary control of muscle glycolysis is the energy charge of the cell*—the ratio of ATP to AMP. Let us examine how each of the key regulatory enzymes responds to changes in the amounts of ATP and AMP present in the cell.

Phosphofructokinase. *Phosphofructokinase is the most important control site in the mammalian glycolytic pathway* (Figure 16.16). High levels of ATP allosterically inhibit the enzyme (a 340-kDa tetramer). ATP binds to a specific regulatory site that is distinct from the catalytic site. The binding of ATP lowers the enzyme's affinity for fructose 6-phosphate. Thus, a high concentration of ATP converts the hyperbolic binding curve of fructose 6-phosphate into a sigmoidal one (Figure 16.17). AMP reverses the inhibitory action of ATP, and so *the activity of the enzyme increases when the ATP/AMP ratio is lowered.* In other words, *glycolysis is stimulated as the energy charge falls.* A decrease in pH also inhibits phosphofructokinase

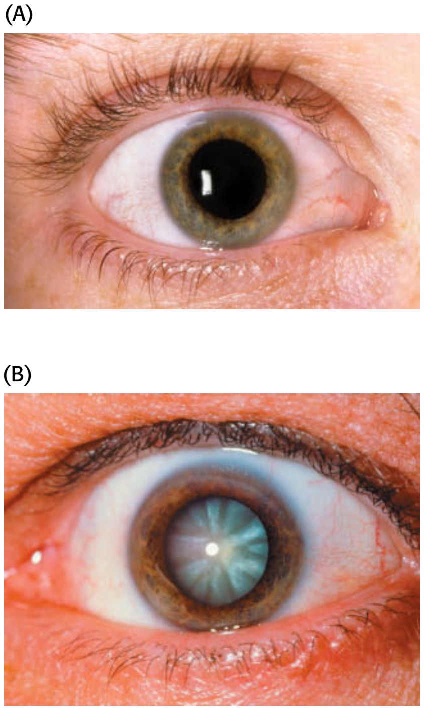


FIGURE 16.15 Cataracts are evident as the clouding of the lens. (A) A healthy eye. (B) An eye with a cataract. [(A) © Imageafter; (B) SPL/Photo Researchers.]

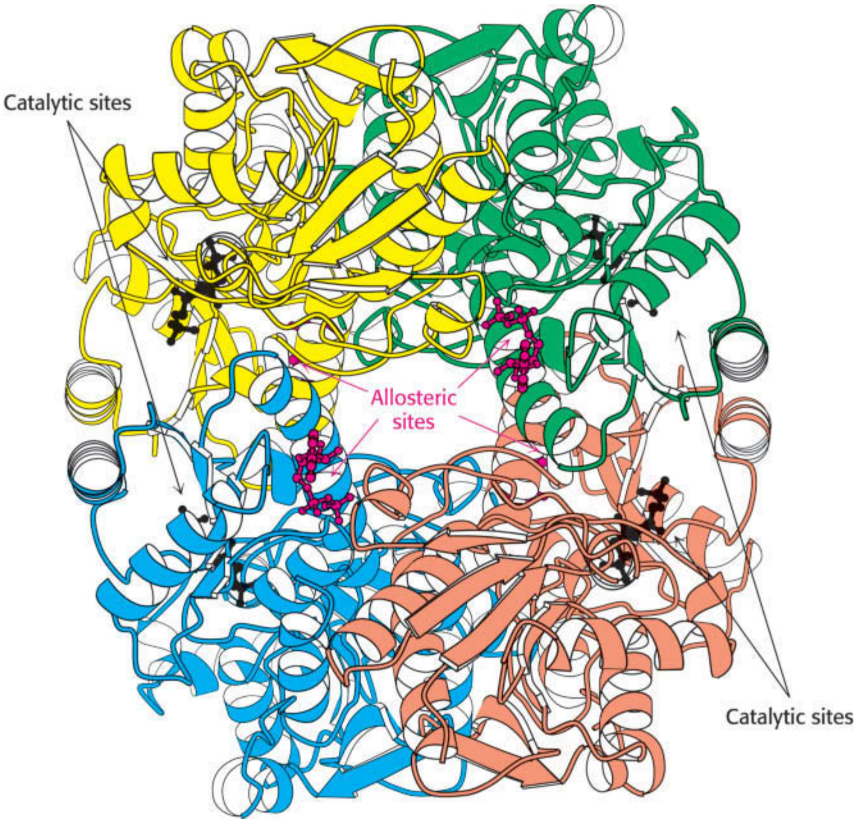


FIGURE 16.16 Structure of phosphofructokinase. The structure of phosphofructokinase from *E. coli* comprises a tetramer of four identical subunits. *Notice* the separation of the catalytic and allosteric sites. Each subunit of the human liver enzyme consists of two domains that are similar to the *E. coli* enzyme. [Drawn from 1PFK.pdb.]

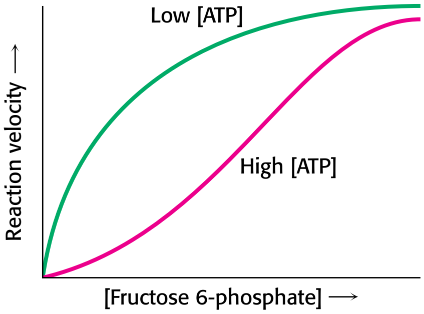


FIGURE 16.17 Allosteric regulation of phosphofructokinase. A high level of ATP inhibits the enzyme by decreasing its affinity for fructose 6-phosphate.

activity by augmenting the inhibitory effect of ATP. The pH might fall when fast twitch muscle is functioning anaerobically, producing excessive quantities of lactic acid. The inhibitory effect protects the muscle from damage that would result from the accumulation of too much acid.

Why is AMP and not ADP the positive regulator of phosphofructokinase? When ATP is being utilized rapidly, the enzyme *adenylate kinase* (Section 9.4) can form ATP from ADP by the following reaction:

$$\text{ADP} + \text{ADP} \rightleftharpoons \text{ATP} + \text{AMP}$$

Thus, some ATP is salvaged from ADP, and AMP becomes the signal for the low-energy state. Moreover, the use of AMP as an allosteric regulator provides an especially sensitive control. We can understand why by considering, first, that the total adenylate pool ([ATP], [ADP], [AMP]) in a cell is constant over the short term and, second, that the concentration of ATP is greater than that of ADP and the concentration of ADP is, in turn, greater than that of AMP. Consequently, small-percentage changes in [ATP] result in larger-percentage changes in the concentrations of the other adenylate nucleotides. This magnification of small changes in [ATP] to larger changes in [AMP] leads to tighter control by increasing the range of sensitivity of phosphofructokinase (Problem 46).

Hexokinase. Phosphofructokinase is the most prominent regulatory enzyme in glycolysis, but it is not the only one. Hexokinase, the enzyme catalyzing the first step of glycolysis, is inhibited by its product, glucose 6-phosphate. High concentrations of this molecule signal that the cell no longer requires glucose for energy or for the synthesis of glycogen, a storage form of glucose

(Chapter 21), and the glucose will be left in the blood. A rise in glucose 6-phosphate concentration is a means by which phosphofructokinase communicates with hexokinase. When phosphofructokinase is inactive, the concentration of fructose 6-phosphate rises. In turn, the level of glucose 6-phosphate rises because it is in equilibrium with fructose 6-phosphate. Hence, *the inhibition of phosphofructokinase leads to the inhibition of hexokinase.*

Why is phosphofructokinase rather than hexokinase the pacemaker of glycolysis? The reason becomes evident on noting that glucose 6-phosphate is not solely a glycolytic intermediate. In muscle, glucose 6-phosphate can also be converted into glycogen. The first irreversible reaction unique to the glycolytic pathway, the *committed step* (Section 10.1), is the phosphorylation of fructose 6-phosphate to fructose 1,6-bisphosphate. Thus, it is highly appropriate for phosphofructokinase to be the primary control site in glycolysis. In general, *the enzyme catalyzing the committed step in a metabolic sequence is the most important control element in the pathway.*

Pyruvate kinase. Pyruvate kinase, the enzyme catalyzing the third irreversible step in glycolysis, controls the outflow from this pathway. This final step yields ATP and pyruvate, a central metabolic intermediate that can be oxidized further or used as a building block. ATP allosterically inhibits pyruvate kinase to slow glycolysis when the energy charge is high. When the pace of glycolysis increases, fructose 1,6-bisphosphate, the product of the preceding irreversible step in glycolysis, activates the kinase to enable it to keep pace with the oncoming high flux of intermediates. A summary of the regulation of glycolysis in resting and active muscle is shown in Figure 16.18.

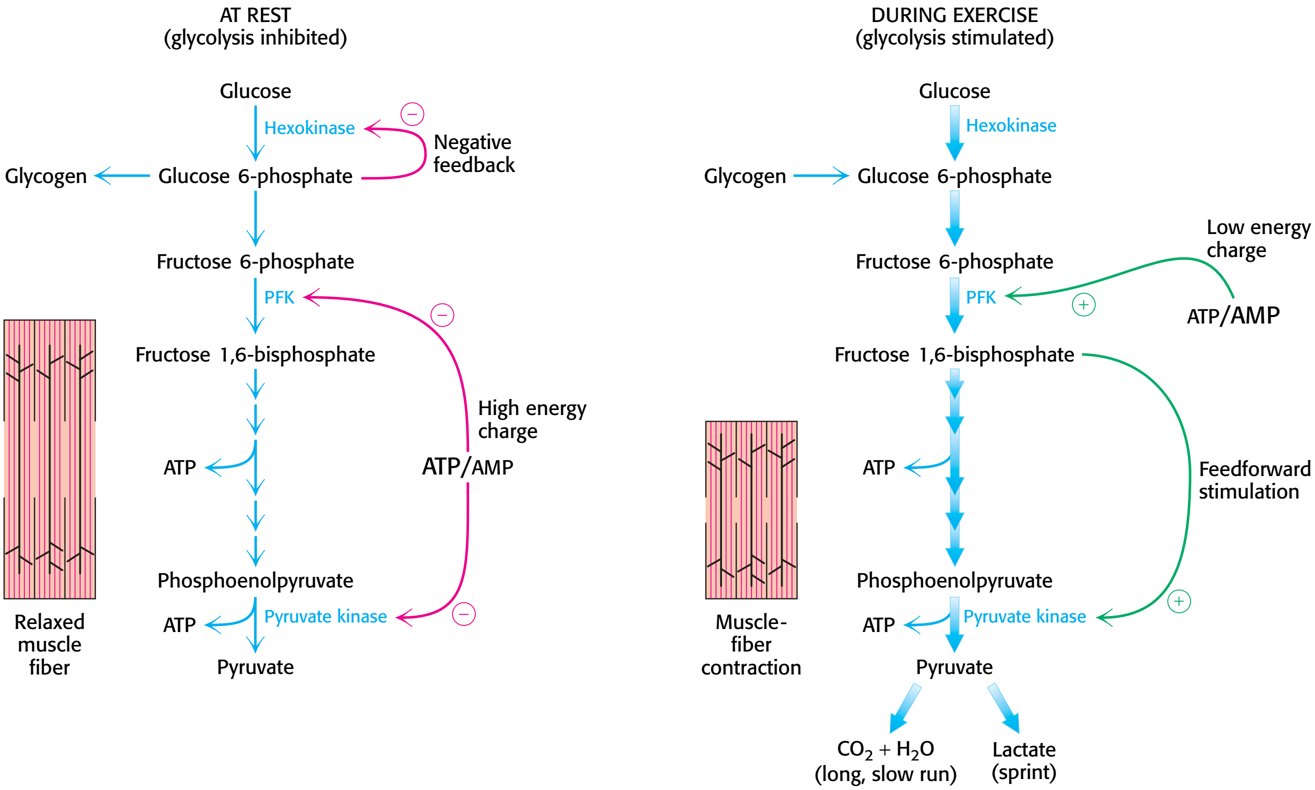

AT REST
(glycolysis inhibited)
Glucose
Hexokinase
−
Negative feedback
Glycogen
Glucose 6-phosphate
Fructose 6-phosphate
PFK
−
Fructose 1,6-bisphosphate
High energy charge
ATP
ATP/AMP
Phosphoenolpyruvate
ATP
Pyruvate kinase
−
Pyruvate
Relaxed muscle fiber
DURING EXERCISE
(glycolysis stimulated)
Glucose
Hexokinase
Glycogen
Glucose 6-phosphate
Low energy charge
Fructose 6-phosphate
PFK
+
ATP/AMP
Fructose 1,6-bisphosphate
ATP
Feedforward stimulation
Phosphoenolpyruvate
+
ATP
Pyruvate kinase
Muscle-fiber contraction
Pyruvate
CO₂ + H₂O
(long, slow run)
Lactate
(sprint)

FIGURE 16.18 **Regulation of glycolysis in muscle.** At rest (left), glycolysis is not very active (thin arrows). The high concentration of ATP inhibits phosphofructokinase (PFK), pyruvate kinase, and hexokinase. Glucose 6-phosphate is converted into glycogen (Chapter 21). During exercise (right), the decrease in the ATP/AMP ratio resulting from muscle contraction activates phosphofructokinase and hence glycolysis. The flux down the pathway is increased, as represented by the thick arrows.

The regulation of glycolysis in the liver illustrates the biochemical versatility of the liver

The liver has more diverse biochemical functions than does muscle. Significantly, the liver maintains blood-glucose levels: it stores glucose as glycogen when glucose is plentiful, and it releases glucose when supplies are low. It also uses glucose to generate reducing power for biosynthesis (Section 20.3) as well as to synthesize a host of biochemicals. So, although the liver has many of the regulatory features of muscle glycolysis, the regulation of glycolysis in the liver is more complex.

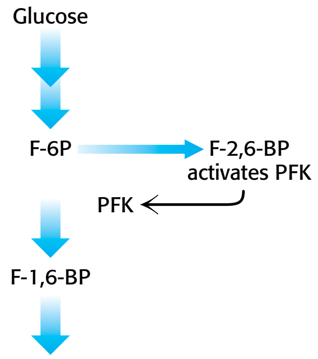
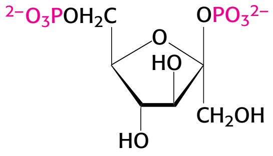


FIGURE 16.19 Regulation of phosphofructokinase by fructose 2,6-bisphosphate. In high concentrations, fructose 6-phosphate (F-6P) activates the enzyme phosphofructokinase (PFK) through an intermediary, fructose 2,6-bisphosphate (F-2,6-BP).

Phosphofructokinase. Liver phosphofructokinase can be regulated by ATP as in muscle, but such regulation is not as important since the liver does not experience the sudden ATP needs that a contracting muscle does. Likewise, low pH is not an important metabolic signal for the liver enzyme, because lactate is not normally produced in the liver. Indeed, as we will see, lactate is converted into glucose in the liver.

Glycolysis in the liver furnishes carbon skeletons for biosyntheses, and so a signal indicating whether building blocks are abundant or scarce should also regulate phosphofructokinase. In the liver, *phosphofructokinase is inhibited by citrate,* an early intermediate in the citric acid cycle (Chapter 17). A high level of citrate in the cytoplasm means that biosynthetic precursors are abundant, and so there is no need to degrade additional glucose for this purpose. Citrate inhibits phosphofructokinase by enhancing the inhibitory effect of ATP.

The key means by which glycolysis in the liver responds to changes in blood glucose is through the signal molecule *fructose 2,6-bisphosphate* (F-2,6-BP), a potent activator of phosphofructokinase (Figure 16.19). In the liver, the concentration of fructose 6-phosphate rises when blood-glucose concentration is high, and the abundance of fructose 6-phosphate accelerates the synthesis of F-2,6-BP (Figure 16.20). Hence, *an abundance of fructose 6-phosphate leads to a higher concentration of F-2,6-BP.* The binding of fructose 2,6-bisphosphate increases the affinity of phosphofructokinase for fructose 6-phosphate and diminishes the inhibitory effect of ATP. Glycolysis is thus accelerated when glucose is abundant. Such a process is called *feedforward stimulation.* We will turn to the synthesis and degradation of this important regulatory molecule after we have considered gluconeogenesis.



Fructose 2,6-bisphosphate (F-2,6-BP)

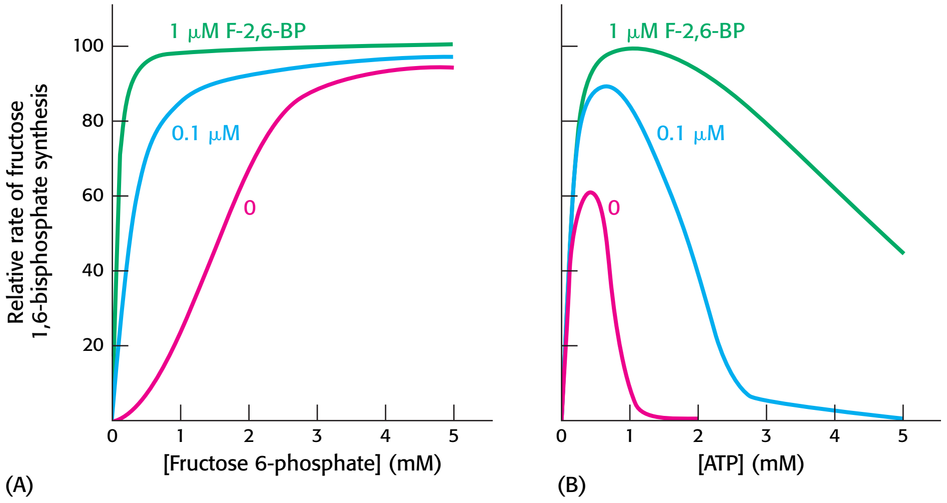


FIGURE 16.20 Activation of phosphofructokinase by fructose 2,6-bisphosphate. (A) The sigmoidal dependence of velocity on substrate concentration becomes hyperbolic in the presence of 1 μM fructose 2,6-bisphosphate. (B) ATP, acting as a substrate, initially stimulates the reaction. As the concentration of ATP increases, it acts as an allosteric inhibitor. The inhibitory effect of ATP is reversed by fructose 2,6-bisphosphate. [Data from E. Van Schaftingen, M. F. Jett, L. Hue, and H. G. Hers, *Proc. Natl. Acad. Sci. U. S. A.* 78:3483–3486, 1981.]

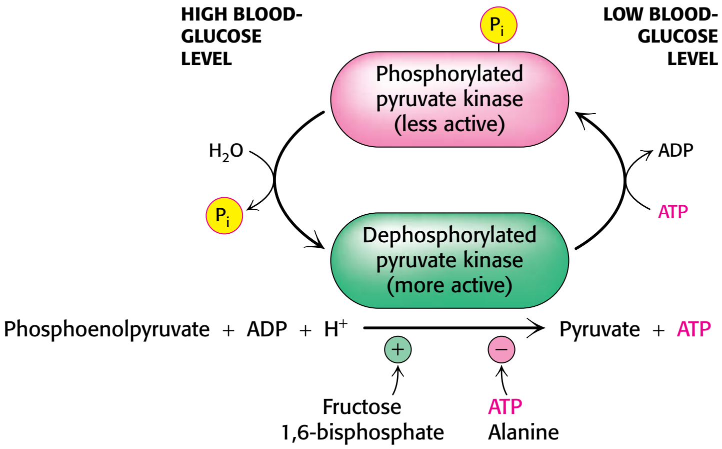

HIGH BLOOD-GLUCOSE LEVEL
LOW BLOOD-GLUCOSE LEVEL
P_i
Phosphorylated pyruvate kinase (less active)
H₂O
P_i
ADP
ATP
Dephosphorylated pyruvate kinase (more active)
Phosphoenolpyruvate + ADP + H+
Pyruvate + ATP
+
−
Fructose 1,6-bisphosphate
ATP
Alanine

FIGURE 16.21 Control of the catalytic activity of pyruvate kinase. Pyruvate kinase is regulated by allosteric effectors and covalent modification. Fructose 1,6-bisphosphate allosterically stimulates the enzyme, while ATP and alanine are allosteric inhibitors. Glucagon, secreted in response to low blood glucose, promotes phosphorylation and inhibition of the enzyme. When blood glucose levels are adequate, the enzyme is dephosphorylated and activated.

Hexokinase and glucokinase. The hexokinase reaction in the liver is controlled as in the muscle. However, the liver, in keeping with its role as monitor of blood-glucose levels, possesses another specialized isozyme of hexokinase, called *glucokinase,* which is not inhibited by glucose 6-phosphate. The role of glucokinase is to provide glucose 6-phosphate for the synthesis of glycogen and for the formation of fatty acids (Section 22.1). Remarkably, glucokinase displays the sigmoidal kinetics characteristic of an allosteric enzyme even though it functions as a monomer. Glucokinase phosphorylates glucose only when glucose is abundant because the affinity of glucokinase for glucose is about 50-fold lower than that of hexokinase. Moreover, when glucose concentration is low, glucokinase is inhibited by the liver-specific glucokinase regulatory protein. The low affinity of glucokinase for glucose gives the brain and muscles first call on glucose when its supply is limited, and it ensures that glucose will not be wasted when it is abundant. Drugs that activate liver glucokinase are being evaluated as a treatment for type 2 or insulin-insensitive diabetes. Glucokinase is also present in the β cells of the pancreas, where the increased formation of glucose 6-phosphate by glucokinase when blood-glucose levels are elevated leads to the secretion of the hormone insulin. Insulin signals the need to remove glucose from the blood for storage as glycogen or conversion into fat.

Pyruvate kinase. Several isozymic forms of pyruvate kinase (a tetramer of 57-kDa subunits) encoded by different genes are present in mammals: the L type predominates in the liver, and the M type in muscle and the brain. The L and M forms of pyruvate kinase have many properties in common. Indeed, the liver enzyme behaves much like the muscle enzyme with regard to allosteric regulation except that the liver enzyme is also inhibited by alanine (synthesized in one step from pyruvate), a signal that building-blocks are available. Moreover, the isozymic forms differ in their susceptibility to covalent modification. The catalytic properties of the L form—but not of the M form—are also controlled by reversible phosphorylation (Figure 16.21). When the blood-glucose level is low, the glucagon-triggered cyclic AMP cascade (p. 483) leads to the phosphorylation of pyruvate kinase, which diminishes its activity. This hormone-triggered phosphorylation prevents the liver from consuming glucose when it is more urgently needed by the brain and muscle. We see here a clear-cut example of how isoenzymes contribute to the metabolic diversity of different organs. We will return to the control of glycolysis after considering gluconeogenesis.

A family of transporters enables glucose to enter and leave animal cells

Several glucose transporters mediate the thermodynamically downhill movement of glucose across the plasma membranes of animal cells. Each member of this protein family, named GLUT1 to GLUT5, consists of a single polypeptide chain about 500 residues long (Table 16.4). Each glucose transporter has a 12-transmembrane-helix structure similar to that of lactose permease (Section 13.3).

TABLE 16.4 Family of glucose transporters

Name	Tissue location	K_M	Comments
GLUT1	All mammalian tissues	1 mM	Basal glucose uptake
GLUT2	Liver and pancreatic β cells	15−20 mM	In the pancreas, plays a role in the regulation of insulin In the liver, removes excess glucose from the blood
GLUT3	All mammalian tissues	1 mM	Basal glucose uptake
GLUT4	Muscle and fat cells	5 mM	Amount in muscle plasma membrane increases with endurance training
GLUT5	Small intestine	—	Primarily a fructose transporter

The members of this family have distinctive roles:

1. GLUT1 and GLUT3, present in nearly all mammalian cells, are responsible for basal glucose uptake. Their K_M value for glucose is about 1 mM, significantly less than the normal serum-glucose level, which typically ranges from 4 mM to 8 mM. Hence, GLUT1 and GLUT3 continually transport glucose into cells at an essentially constant rate.

2. GLUT2, present in liver and pancreatic β cells, is distinctive in having a very high K_M value for glucose (15–20 mM). Hence, glucose enters these tissues at a biologically significant rate only when there is much glucose in the blood. The pancreas can sense the glucose level and accordingly adjust the rate of insulin secretion. The high K_M value of GLUT2 also ensures that glucose rapidly enters liver cells only in times of plenty.

3. GLUT4, which has a K_M value of 5 mM, transports glucose into muscle and fat cells. The number of GLUT4 transporters in the plasma membrane increases rapidly in the presence of insulin, which signals the fed state. Hence, insulin promotes the uptake of glucose by muscle and fat. Endurance exercise training increases the amount of this transporter present in muscle membranes.

4. GLUT5, present in the small intestine, functions primarily as a fructose transporter.

Aerobic glycolysis is a property of rapidly growing cells.

Tumors have been known for decades to display enhanced rates of glucose uptake and glycolysis. Indeed, rapidly growing tumor cells will metabolize glucose to lactate even in the presence of oxygen, a process called *aerobic glycolysis* or the *Warburg effect*, after Otto Warburg, the biochemist who first noted this characteristic of cancer cells in the 1920s. In fact, tumors with a high glucose uptake are particularly aggressive, and the cancer is likely to have a poor prognosis. A nonmetabolizable glucose analog, 2-^{18}F-2-D-deoxyglucose, detectable by a combination of positron emission tomography (PET) and computer-aided tomography (CAT), easily visualizes tumors and allows monitoring of the effectiveness of treatment (Figure 16.22).

What selective advantage does aerobic glycolysis offer the tumor over the energetically more efficient oxidative phosphorylation? Research is being actively pursued to answer the question, but we can speculate on the benefits. First, aerobic glycolysis generates lactic acid that is then secreted. Acidification of the tumor environment has been shown to facilitate tumor invasion and inhibit the immune system from attacking the tumor. However, even leukemias perform aerobic glycolysis, and leukemia is not

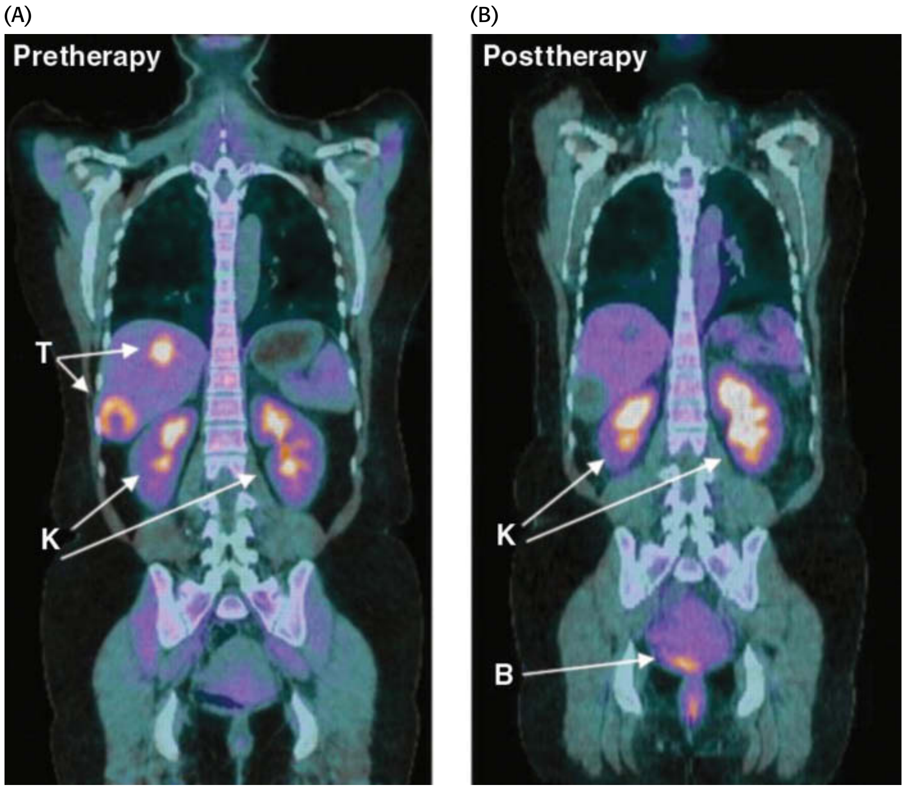


FIGURE 16.22 Tumors can be visualized with 2-^{18}F-2-D-deoxyglucose (FDG) and positron emission tomography. (A) A nonmetabolizable glucose analog infused into a patient and detected by a combination of positron emission and computer-aided tomography reveals the presence of a malignant tumor (T). (B) After 4 weeks of treatment with a tyrosine kinase inhibitor (Section 14.5), the tumor shows no uptake of FDG, indicating decreased metabolism. Excess FDG, which is excreted in the urine, also visualizes the kidneys (K) and bladder (B). [Images courtesy of A. D. Van den Abbeele, Dana-Farber Cancer Institute, Boston.]

an invasive cancer. Second, and perhaps more importantly, the increased uptake of glucose and formation of glucose 6-phosphate provides substrates for another metabolic pathway, the pentose phosphate pathway (Chapter 20), that generates biosynthetic reducing power. Moreover, the pentose phosphate pathway, in cooperation with glycolysis, produces precursors for biomolecules necessary for growth, such as nucleotides. Finally, cancer cells grow more rapidly than the blood vessels that nourish them; thus, as solid tumors grow, the oxygen concentration in their environment falls. In other words, they begin to experience *hypoxia,* a deficiency of oxygen. The use of aerobic glycolysis reduces the dependence of cell growth on oxygen. Not all of the precursor needs are met by enhanced glucose metabolism. Cancer cells also require glutamine, which is channeled into the mitochondria to replenish citric acid cycle components used for biosynthesis.

What biochemical alterations facilitate the switch to aerobic glycolysis? Again, the answers are not complete, but changes in gene expression of isozymic forms of two glycolytic enzymes may be crucial. Tumor cells express an isozyme of hexokinase that binds to mitochondria. There, the enzyme has ready access to any ATP generated by oxidative phosphorylation and is not susceptible to feedback inhibition by its product, glucose 6-phosphate. More importantly, an embryonic isozyme of pyruvate kinase, pyruvate kinase M, is also expressed. Remarkably, this isozyme has a lower catalytic rate than normal pyruvate kinase and creates a bottleneck, allowing the use of glycolytic intermediates for biosynthetic processes required for

TABLE 16.5 Proteins in glucose metabolism encoded by genes regulated by hypoxia-inducible factor

GLUT1
GLUT3
Hexokinase
Phosphofructokinase
Aldolase
Glyceraldehyde 3-phosphate dehydrogenase
Phosphoglycerate kinase
Enolase
Pyruvate kinase

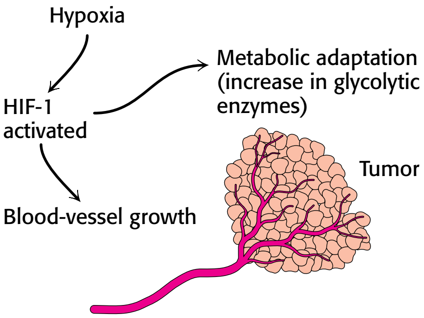


FIGURE 16.23 Alteration of gene expression in tumors owing to hypoxia. The hypoxic conditions inside a tumor mass lead to the activation of the hypoxia-inducible transcription factor (HIF-1), which induces metabolic adaptation (an increase in glycolytic enzymes) and activates angiogenic factors that stimulate the growth of new blood vessels. [Information from C. V. Dang and G. L. Semenza, *Trends Biochem. Sci.* 24:68–72, 1999.]

cell proliferation. The need for biosynthetic precursors is greater than the need for ATP, suggesting that even glycolysis at a reduced rate produces sufficient ATP to allow cell proliferation. Although originally observed in cancer cells, the Warburg effect is also seen in noncancerous, rapidly dividing cells.

Cancer and endurance training affect glycolysis in a similar fashion

The hypoxia that some tumors experience with rapid growth activates a transcription factor, *hypoxia-inducible transcription factor* (HIF-1). HIF-1 increases the expression of most glycolytic enzymes and the glucose transporters GLUT1 and GLUT3 (Table 16.5). These adaptations by the cancer cells enable a tumor to survive until blood vessels can grow. HIF-1 also increases the expression of signal molecules, such as vascular endothelial growth factor (VEGF), that facilitate the growth of blood vessels that will provide nutrients to the cells (Figure 16.23). Without new blood vessels, a tumor would cease to grow and either die or remain harmlessly small. Efforts are underway to develop drugs that inhibit the growth of blood vessels in tumors.

Interestingly, anaerobic exercise training—forcing muscles to rely on lactic acid fermentation for ATP production—also activates HIF-1, producing the same effects as those seen in the tumor—enhanced ability to generate ATP anaerobically and a stimulation of blood-vessel growth. These biochemical effects account for the improved athletic performance that results from training and demonstrate how behavior can affect biochemistry. Other signals from sustained muscle contraction trigger muscle mitochondrial biogenesis, allowing for more efficient aerobic energy generation and forestalling the need to resort to lactic acid fermentation for ATP synthesis (Chapter 27).

16.3 Glucose Can Be Synthesized from Noncarbohydrate Precursors

We now turn to the *synthesis of glucose from noncarbohydrate precursors,* a process called *gluconeogenesis.* Maintaining levels of glucose is important because the brain depends on glucose as its primary fuel and red blood cells use glucose as their only fuel. The daily glucose requirement of the brain in a typical adult human being is about 120 g, which accounts for most of the 160 g of glucose needed daily by the whole body. The amount of glucose present in body fluids is about 20 g, and that readily available from glycogen is approximately 190 g. Thus, the direct glucose reserves are sufficient to meet glucose needs for about a day. Gluconeogenesis is especially important during a longer period of fasting or starvation (Section 27.5).

The gluconeogenic pathway converts pyruvate into glucose. Noncarbohydrate precursors of glucose are first converted into pyruvate or enter the pathway at later intermediates such as oxaloacetate and dihydroxyacetone phosphate (Figure 16.24). The major noncarbohydrate precursors are *lactate, amino acids,* and *glycerol.* Lactate is formed by active skeletal muscle when the rate of glycolysis exceeds the rate of oxidative metabolism. Lactate is readily converted into pyruvate by the action of lactate dehydrogenase (p. 464). Amino acids are derived from proteins in the diet and, during starvation, from the breakdown of proteins in skeletal muscle (Section 23.1). The hydrolysis of triacylglycerols (Section 22.2) in fat cells yields glycerol and

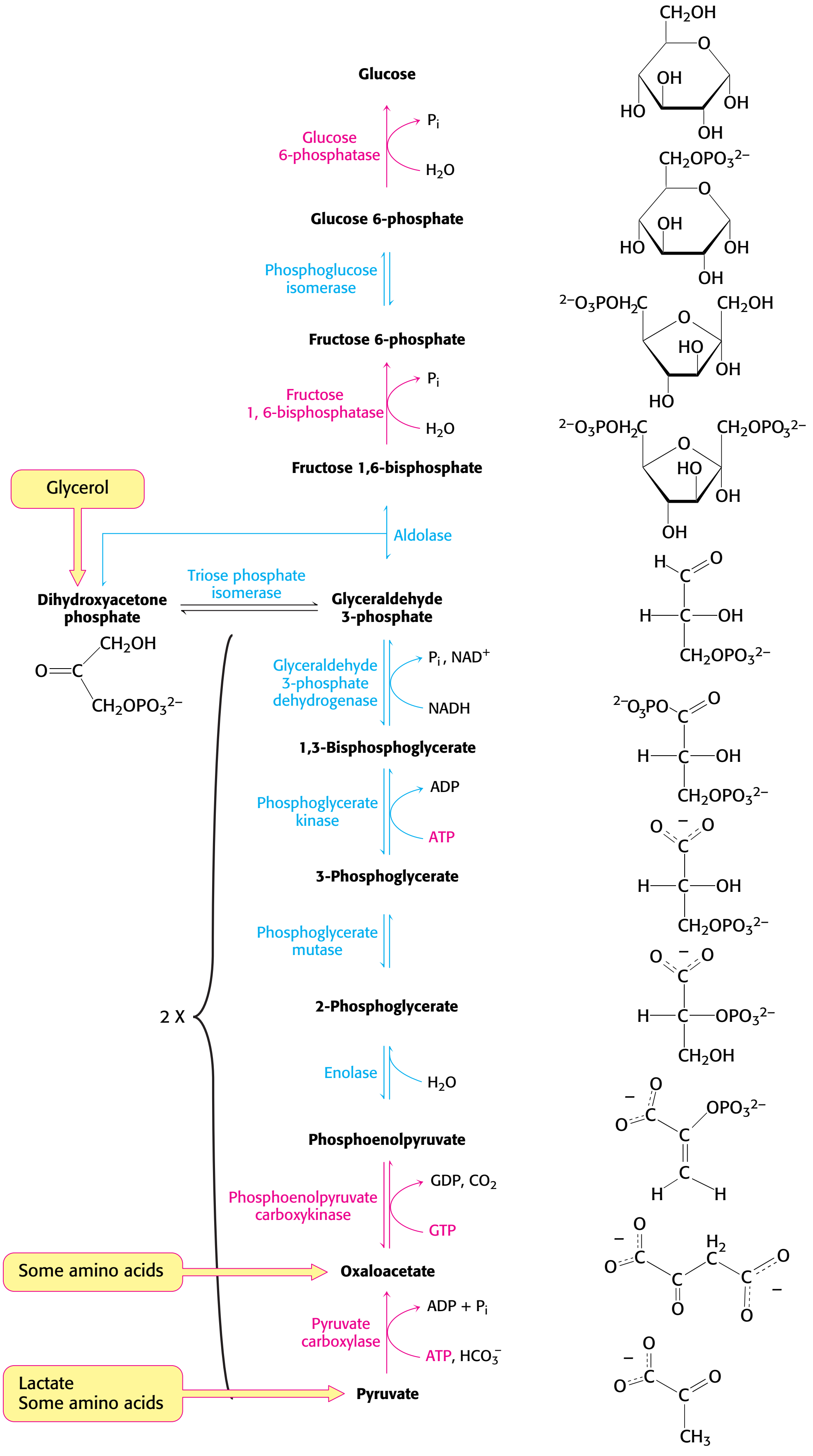
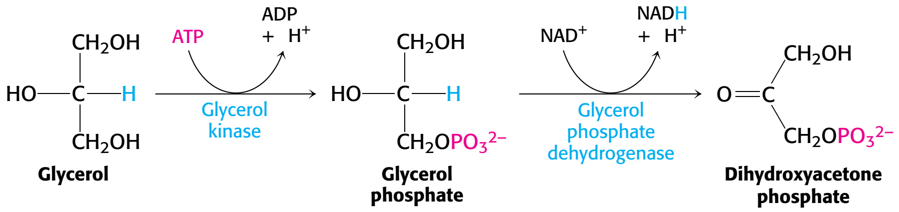

Glucose
Glucose 6-phosphatase
P_i
H_2O
Glucose 6-phosphate
Phosphoglucose isomerase
Fructose 6-phosphate
Fructose 1, 6-bisphosphatase
P_i
H_2O
Fructose 1,6-bisphosphate
Glycerol
Aldolase
Dihydroxyacetone phosphate
Triose phosphate isomerase
Glyceraldehyde 3-phosphate
Glyceraldehyde 3-phosphate dehydrogenase
P_i, NAD^+
NADH
1,3-Bisphosphoglycerate
Phosphoglycerate kinase
ADP
ATP
3-Phosphoglycerate
Phosphoglycerate mutase
2-Phosphoglycerate
2 X
Enolase
H_2O
Phosphoenolpyruvate
Phosphoenolpyruvate carboxykinase
GDP, CO_2
GTP
Some amino acids
Oxaloacetate
Pyruvate carboxylase
ADP + P_i
ATP, HCO_3^-
Lactate
Some amino acids
Pyruvate
CH_2OH
$CH_2OPO_3^{2-}$
$^{2-}O_3POH_2C$
$CH_2OPO_3^{2-}$
OPO_3^{2-}
CH_3

FIGURE 16.24 Pathway of gluconeogenesis. The reactions and enzymes unique to gluconeogenesis are shown in red. The other reactions are common to glycolysis. The enzymes for gluconeogenesis are located in the cytoplasm, except for pyruvate carboxylase (in the mitochondria) and glucose 6-phosphatase (membrane bound in the endoplasmic reticulum). The entry points for lactate, glycerol, and amino acids are shown.

fatty acids. Glycerol is a precursor of glucose, but animals cannot convert fatty acids into glucose, for reasons that will be given later. Glycerol may enter either the gluconeogenic or the glycolytic pathway at dihydroxyacetone phosphate.


CH2OH
HO—C—H
CH2OH
Glycerol
ATP
ADP + H+
Glycerol kinase
CH2OH
HO—C—H
CH2OPO3 2−
Glycerol phosphate
NAD+
NADH + H+
Glycerol phosphate dehydrogenase
CH2OH
O=C
CH2OPO3 2−
Dihydroxyacetone phosphate

The major site of gluconeogenesis is the *liver,* with a small amount also taking place in the *kidney.* Little gluconeogenesis takes place in the brain, skeletal muscle, or heart muscle. Rather, *gluconeogenesis in the liver and kidney helps to maintain the glucose level in the blood so that the brain and muscle can extract sufficient glucose from it to meet their metabolic demands.*

Gluconeogenesis is not a reversal of glycolysis

In glycolysis, glucose is converted into pyruvate; in gluconeogenesis, pyruvate is converted into glucose. However, *gluconeogenesis is not a reversal of glycolysis.* Several reactions must differ because the equilibrium of glycolysis lies far on the side of pyruvate formation. The actual free energy change for the formation of pyruvate from glucose is about -90 kJ mol^{-1} (-22 kcal mol^{-1}) under typical cellular conditions. Most of the decrease in free energy in glycolysis takes place in the three essentially irreversible steps catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase.

$$\text{Glucose} + \text{ATP} \xrightarrow{\text{Hexokinase}} \text{Glucose 6-phosphate} + \text{ADP}$$
$$\Delta G = -33 \text{ kJ mol}^{-1} \ (-8.0 \text{ kcal mol}^{-1})$$

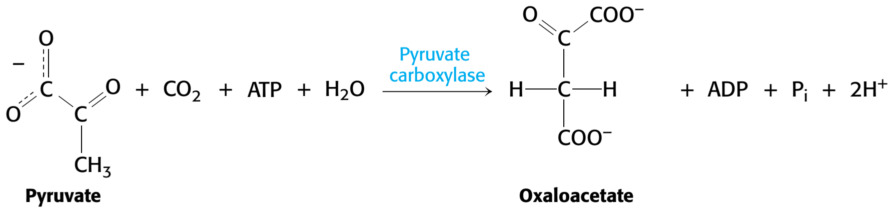
$$\text{Fructose 6-phosphate} + \text{ATP} \xrightarrow{\text{Phosphofructokinase}}$$
$$\text{fructose 1,6-bisphosphate} + \text{ADP}$$
$$\Delta G = -22 \text{ kJ mol}^{-1} \ (-5.3 \text{ kcal mol}^{-1})$$

$$\text{Phosphoenolpyruvate} + \text{ADP} \xrightarrow{\text{Pyruvate kinase}} \text{pyruvate} + \text{ATP}$$
$$\Delta G = -17 \text{ kJ mol}^{-1} \ (-4.0 \text{ kcal mol}^{-1})$$

In gluconeogenesis, these virtually irreversible reactions of glycolysis must be bypassed.

The conversion of pyruvate into phosphoenolpyruvate begins with the formation of oxaloacetate

The first step in gluconeogenesis is the carboxylation of pyruvate to form oxaloacetate at the expense of a molecule of ATP, a reaction catalyzed by *pyruvate carboxylase*. This reaction occurs in the mitochondria.


Pyruvate
+ CO2 + ATP + H2O
Pyruvate carboxylase
Oxaloacetate
+ ADP + Pi + 2H+

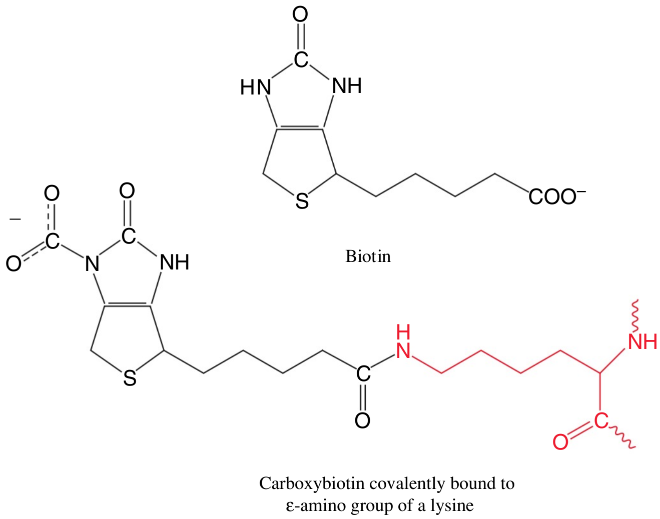


FIGURE 16.25 Structure of biotin and carboxybiotin.

Pyruvate carboxylase requires *biotin,* a covalently attached prosthetic group, which serves as the *carrier of activated CO_2.* The carboxylate group of biotin is linked to the ε-amino group of a specific lysine residue by an amide bond (Figure 16.25). Recall that, in aqueous solutions, CO_2 exists primarily as HCO_3^- with the aid of carbonic anhydrase (Section 9.2).

The carboxylation of pyruvate takes place in three stages:

$$HCO_3^- + ATP \rightleftharpoons HOCO_2\text{-}PO_3^{2-} + ADP$$

$$\text{Biotin–enzyme} + HOCO_2\text{-}PO_3^{2-} \rightleftharpoons CO_2 - \text{biotin–enzyme} + P_i$$

$$CO_2 - \text{biotin–enzyme} + \text{pyruvate} \rightleftharpoons \text{biotin–enzyme} + \text{oxaloacetate}$$

Pyruvate carboxylase functions as a tetramer composed of four identical subunits, and each subunit consists of four domains (Figure 16.26). The biotin carboxylase domain (BC) catalyzes the formation of carboxyphosphate and the subsequent attachment of CO_2 to the second domain, the biotin carboxyl carrier protein (BCCP), the site of the covalently attached biotin. Once bound to CO_2, BCCP leaves the biotin carboxylase active site and swings almost the entire length of the subunit (≈75Å) to the active site of the pyruvate carboxylase domain (PC), which transfers the CO_2 to pyruvate to form oxaloacetate. BCCP in one subunit interacts with the active sites on an adjacent subunit. The fourth domain (PT) facilitates the formation of the tetramer.

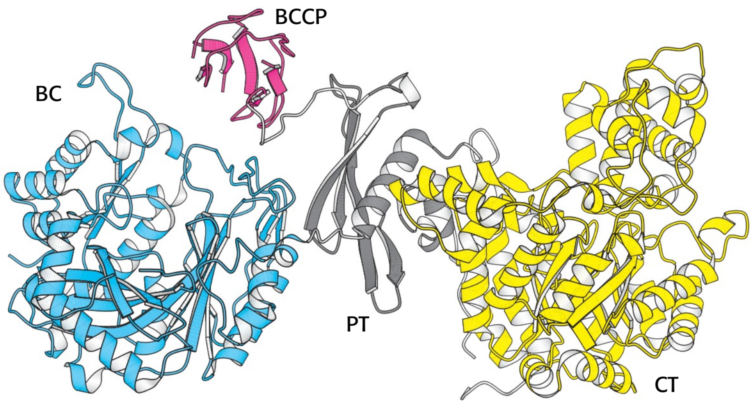


FIGURE 16.26 A subunit of pyruvate carboxylase. Biotin, covalently attached to the BCCP, transports CO_2 from the BC active site to the PC active site of an adjacent subunit. (BC) biotin carboxylase; (BCCP) biotin carboxyl carrier protein; (PC) pyruvate carboxylase; (PT) *p*yruvate carboxylase *t*etramerization domain. [Information from G. Lasso, L.P.C. Yu, D. Gil, S. Xiang, L. Tong, and M. Valle, *Structure* 18:1300–1310, 2010.]

The first partial reaction of pyruvate carboxylase, the formation of carboxybiotin, depends on the presence of acetyl CoA. *Biotin is not carboxylated unless acetyl CoA is bound to the enzyme.* Acetyl CoA has no effect on the second partial reaction. The allosteric activation of pyruvate carboxylase by acetyl CoA is an important physiological control mechanism that will be discussed in Section 17.4.

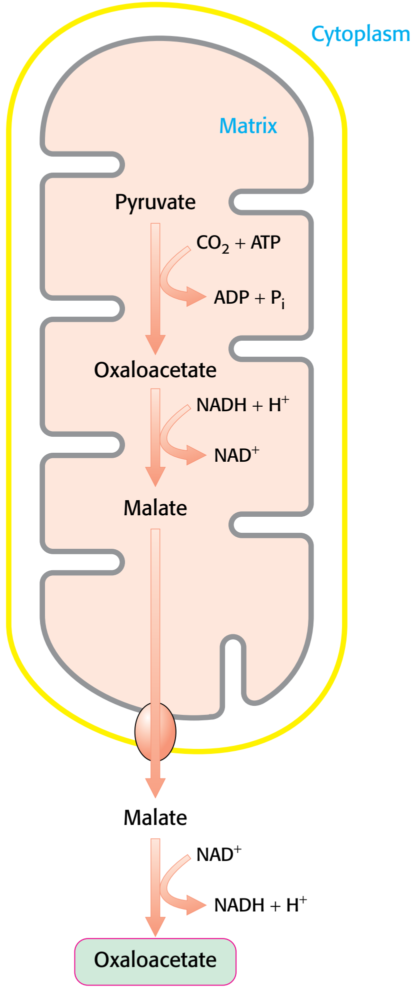
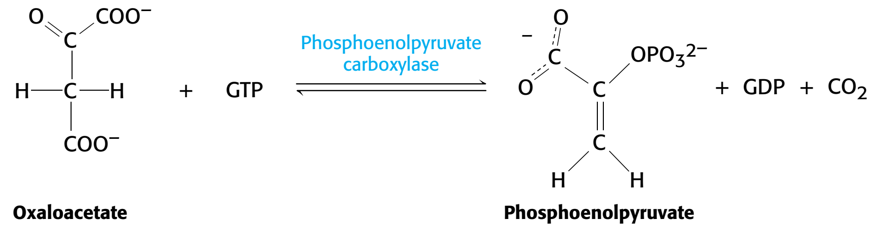


FIGURE 16.27 Compartmental cooperation. Oxaloacetate used in the cytoplasm for gluconeogenesis is formed in the mitochondrial matrix by the carboxylation of pyruvate. Oxaloacetate leaves the mitochondrion by a specific transport system (not shown) in the form of malate, which is reoxidized to oxaloacetate in the cytoplasm.

Oxaloacetate is shuttled into the cytoplasm and converted into phosphoenolpyruvate

Oxaloacetate must thus be transported to the cytoplasm to complete the synthesis of phosphoenolpyruvate. Oxaloacetate is first reduced to malate by malate dehydrogenase. Malate is transported across the mitochondrial membrane and reoxidized to oxaloacetate by a cytoplasmic NAD^+-linked malate dehydrogenase (Figure 16.27). The formation of oxaloacetate from malate also provides NADH for use in subsequent steps in gluconeogenesis. Finally, oxaloacetate is simultaneously *decarboxylated* and *phosphorylated* by *phosphoenolpyruvate carboxykinase* (PEPCK) to generate phosphoenolpyruvate. The phosphoryl donor is GTP. The CO_2 that was added to pyruvate by pyruvate carboxylase comes off in this step.



The sum of the reactions catalyzed by pyruvate carboxylase and phosphoenolpyruvate carboxykinase is

$$\text{Pyruvate} + \text{ATP} + \text{GTP} + H_2O \longrightarrow \text{phosphoenolpyruvate} + \text{ADP} + \text{GDP} + P_i + 2\,H^+$$

This pair of reactions bypasses the irreversible reaction catalyzed by pyruvate kinase in glycolysis.

Why is a carboxylation and a decarboxylation required to form phosphoenolpyruvate from pyruvate? Recall that, in glycolysis, the presence of a phosphoryl group traps the unstable enol isomer of pyruvate as phosphoenolpyruvate (p. 461). However, the addition of a phosphoryl group to pyruvate is a highly unfavorable reaction: the $\Delta G^{\circ\prime}$ of the reverse of the glycolytic reaction catalyzed by pyruvate kinase is $+31$ kJ mol^{-1} ($+7.5$ kcal mol^{-1}). In gluconeogenesis, the use of the carboxylation and decarboxylation steps results in a much more favorable $\Delta G^{\circ\prime}$. The formation of phosphoenolpyruvate from pyruvate in the gluconeogenic pathway has a $\Delta G^{\circ\prime}$ of $+0.8$ kJ mol^{-1} ($+0.2$ kcal mol^{-1}). A molecule of ATP is used to power the addition of a molecule of CO_2 to pyruvate in the carboxylation step. That CO_2 is then removed to power the formation of phosphoenolpyruvate in the decarboxylation step. *Decarboxylations often drive reactions that are otherwise highly endergonic.* This metabolic motif is used in the citric acid cycle (Chapter 17), the pentose phosphate pathway (Chapter 20), and fatty acid synthesis (Section 22.4).

The conversion of fructose 1,6-bisphosphate into fructose 6-phosphate and orthophosphate is an irreversible step

On formation, phosphoenolpyruvate is metabolized by the enzymes of glycolysis but in the reverse direction. These reactions are near equilibrium under intracellular conditions; so, when conditions favor gluconeogenesis, the reverse reactions will take place until the next irreversible step is reached. This step is the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate and P_i.

$$\text{Fructose 1,6-bisphosphate} + H_2O \xrightarrow{\text{Fructose 1,6-bisphosphatase}} \text{fructose 6-phosphate} + P_i$$

The enzyme responsible for this step is *fructose 1,6-bisphosphatase*. Like its glycolytic counterpart, it is an allosteric enzyme that participates in the regulation of gluconeogenesis. We will return to its regulatory properties later in the chapter.

The generation of free glucose is an important control point

The fructose 6-phosphate generated by fructose 1,6-bisphosphatase is readily converted into glucose 6-phosphate. In most tissues, gluconeogenesis ends here. Free glucose is not generated; rather, the glucose 6-phosphate is processed in some other fashion, notably to form glycogen. One advantage to ending gluconeogenesis at glucose 6-phosphate is that, unlike free glucose, the molecule is not transported out of the cell. The enzyme responsible for the conversion of glucose 6-phosphate into glucose, *glucose 6-phosphatase,* is present only in tissues whose metabolic duty is to maintain blood-glucose homeostasis—tissues that release glucose into the blood. These tissues are the liver and to a lesser extent the kidney.

This final step in the generation of glucose does not take place in the cytoplasm. Rather, glucose 6-phosphate is transported into the lumen of the endoplasmic reticulum, where it is hydrolyzed to glucose by glucose 6-phosphatase, which is bound to the membrane (Figure 16.28). Glucose and P_i are then shuttled back to the cytoplasm by a pair of transporters. The glucose transporter in the endoplasmic reticulum membrane is like those found in the plasma membrane.

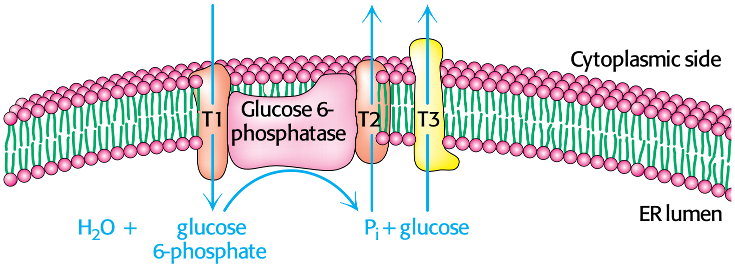


FIGURE 16.28 Generation of glucose from glucose 6-phosphate. Several endoplasmic reticulum (ER) proteins play a role in the generation of glucose from glucose 6-phosphate. T1 transports glucose 6-phosphate into the lumen of the ER, whereas T2 and T3 transport P_i and glucose, respectively, back into the cytoplasm. [Information from A. Buchell and I. D. Waddel, *Biochem. Biophys. Acta* 1092:129–137, 1991.]

Six high-transfer-potential phosphoryl groups are spent in synthesizing glucose from pyruvate

The formation of glucose from pyruvate is energetically unfavorable unless it is coupled to reactions that are favorable. Compare the stoichiometry of gluconeogenesis with that of the reverse of glycolysis.

The stoichiometry of gluconeogenesis is

$$\begin{aligned} &2\,\text{Pyruvate} + 4\,\text{ATP} + 2\,\text{GTP} + 2\,\text{NADH} + 6\,H_2O \rightarrow \\ &\qquad \text{glucose} + 4\,\text{ADP} + 2\,\text{GDP} + 6\,P_i + 2\,\text{NAD}^+ + 2\,H^+ \\ &\qquad \Delta G^{\circ\prime} = -48\ \text{kJ mol}^{-1}\ (-11\ \text{kcal mol}^{-1}) \end{aligned}$$

In contrast, the stoichiometry for the reversal of glycolysis is

$$\begin{aligned} &2\,\text{Pyruvate} + 2\,\text{ATP} + \text{NADH} + 2\,H_2O \rightarrow \\ &\qquad \text{glucose} + 2\,\text{ADP} + 2\,P_i + 2\,\text{NAD}^+ + 2H^+ \\ &\qquad \Delta G^{\circ\prime} = +90\ \text{kJ mol}^{-1}\ (+22\ \text{kcal mol}^{-1}) \end{aligned}$$

Note that *six* nucleoside triphosphate molecules are hydrolyzed to synthesize glucose from pyruvate in gluconeogenesis, whereas only *two* molecules of ATP are generated in glycolysis in the conversion of glucose into pyruvate.

TABLE 16.6 Reactions of gluconeogenesis

Step	Reaction
1	Pyruvate + CO_2 + ATP + $H_2O \longrightarrow$ oxaloacetate + ADP + P_i + $2H^+$
2	Oxaloacetate + GTP $\rightleftharpoons$ phosphoenolpyruvate + GDP + CO_2
3	Phosphoenolpyruvate + $H_2O \rightleftharpoons$ 2-phosphoglycerate
4	2-Phosphoglycerate $\rightleftharpoons$ 3-phosphoglycerate
5	3-Phosphoglycerate + ATP $\rightleftharpoons$ 1,3-bisphosphoglycerate + ADP
6	1,3-Bisphosphoglycerate + NADH + $H^+ \rightleftharpoons$ glyceraldehyde 3-phosphate + NAD^+ + P_i
7	Glyceraldehyde 3-phosphate $\rightleftharpoons$ dihydroxyacetone phosphate
8	Glyceraldehyde 3-phosphate + dihydroxyacetone phosphate $\rightleftharpoons$ fructose 1,6-bisphosphate
9	Fructose 1,6-bisphosphate + $H_2O \longrightarrow$ fructose 6-phosphate + P_i
10	Fructose 6-phosphate $\rightleftharpoons$ glucose 6-phosphate
11	Glucose 6-phosphate + $H_2O \longrightarrow$ + glucose + P_i

Thus, the extra cost of gluconeogenesis is four high-phosphoryl-transfer-potential molecules for each molecule of glucose synthesized from pyruvate. The four additional molecules having high phosphoryl-transfer potential are needed to turn an energetically unfavorable process (the reversal of glycolysis) into a favorable one (gluconeogenesis). Here we have a clear example of the coupling of reactions: NTP hydrolysis is used to power an energetically unfavorable reaction. The reactions of gluconeogenesis are summarized in Table 16.6.

16.4 Gluconeogenesis and Glycolysis Are Reciprocally Regulated

Gluconeogenesis and glycolysis are coordinated so that, within a cell, one pathway is relatively inactive while the other is highly active. If both sets of reactions were highly active at the same time, the net result would be the hydrolysis of four nucleoside triphosphates (two ATP molecules plus two GTP molecules) per reaction cycle. Both glycolysis and gluconeogenesis are highly exergonic under cellular conditions, and so there is no thermodynamic barrier to such simultaneous activity. However, the *amounts* and *activities* of the distinctive enzymes of each pathway are controlled so that both pathways are not highly active at the same time. The rate of glycolysis is also determined by the concentration of glucose, and the rate of gluconeogenesis by the concentrations of lactate and other precursors of glucose. The basic premise of the reciprocal regulation is that, when energy is needed or glycolytic intermediates are needed for biosynthesis, glycolysis will predominate. When there is a surplus of energy and glucose precursors, gluconeogenesis will take over.

Energy charge determines whether glycolysis or gluconeogenesis will be most active

The first important regulation site is the interconversion of fructose 6-phosphate and fructose 1,6-bisphosphate (Figure 16.29). Consider first a situation in which energy is needed. In this case, the concentration of AMP is high. Under this condition, AMP stimulates phosphofructokinase but inhibits fructose 1,6-bisphosphatase. Thus, glycolysis is turned on and gluconeogenesis is inhibited. Conversely, high levels of ATP and citrate indicate that the energy charge is high and that biosynthetic intermediates are abundant. ATP and citrate inhibit phosphofructokinase, whereas citrate activates fructose 1,6-bisphosphatase. Under these conditions, glycolysis is nearly switched off and gluconeogenesis is promoted. Why does citrate take

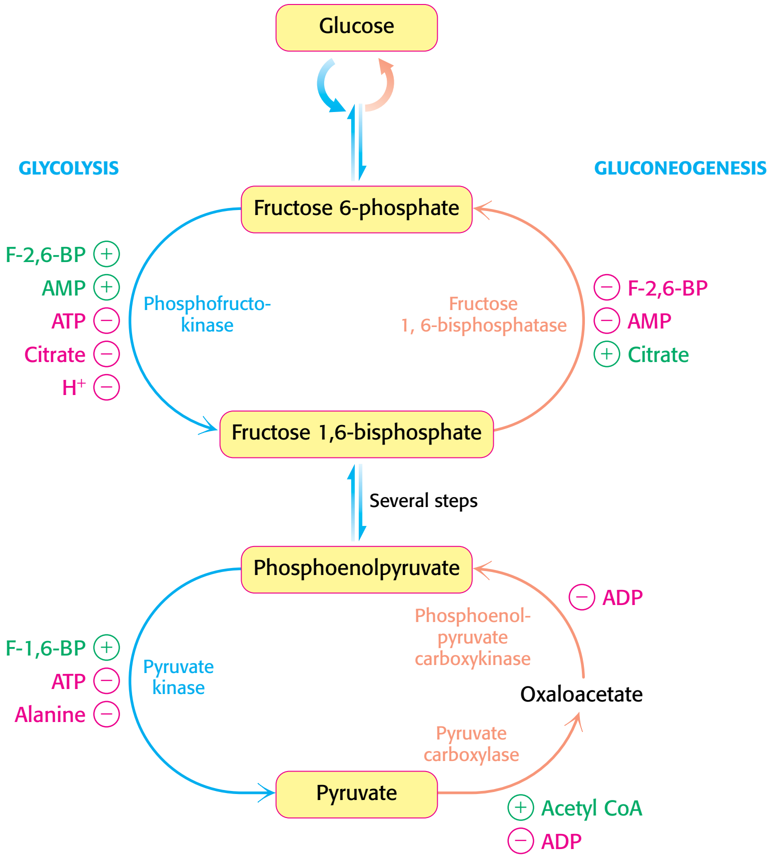

Glucose
GLYCOLYSIS
GLUCONEOGENESIS
Fructose 6-phosphate
F-2,6-BP +
AMP +
ATP −
Citrate −
H+ −
Phosphofructo-kinase
Fructose 1, 6-bisphosphatase
− F-2,6-BP
− AMP
+ Citrate
Fructose 1,6-bisphosphate
Several steps
Phosphoenolpyruvate
− ADP
Phosphoenol-pyruvate carboxykinase
F-1,6-BP +
ATP −
Alanine −
Pyruvate kinase
Oxaloacetate
Pyruvate carboxylase
Pyruvate
+ Acetyl CoA
− ADP

FIGURE 16.29 Reciprocal regulation of gluconeogenesis and glycolysis in the liver. The level of fructose 2,6-bisphosphate is high in the fed state and low in starvation. Another important control is the inhibition of pyruvate kinase by phosphorylation during starvation.

part in this regulatory scheme? As we will see in Chapter 17, citrate reports on the status of the citric acid cycle, the primary pathway for oxidizing fuels in the presence of oxygen. High levels of citrate indicate an energy-rich situation and the presence of precursors for biosynthesis.

Glycolysis and gluconeogenesis are also reciprocally regulated at the interconversion of phosphoenolpyruvate and pyruvate in the liver. The glycolytic enzyme pyruvate kinase is inhibited by allosteric effectors ATP and alanine, which signal that the energy charge is high and that building blocks are abundant. Conversely, pyruvate carboxylase, which catalyzes the first step in gluconeogenesis from pyruvate, is inhibited by ADP. Likewise, ADP inhibits phosphoenolpyruvate carboxykinase. Pyruvate carboxylase is activated by acetyl CoA, which, like citrate, indicates that the citric acid cycle is producing energy and biosynthetic intermediates (Chapter 17). Hence, gluconeogenesis is favored when the cell is rich in biosynthetic precursors and ATP.

The balance between glycolysis and gluconeogenesis in the liver is sensitive to blood-glucose concentration

In the liver, rates of glycolysis and gluconeogenesis are adjusted to maintain blood-glucose levels. *The signal molecule fructose 2,6-bisphosphate strongly stimulates phosphofructokinase* (PFK) *and inhibits fructose 1,6-bisphosphatase.* When blood glucose is low, fructose 2,6-bisphosphate loses a phosphoryl group to form fructose 6-phosphate, which is not an allosteric effector of PFK. How is the concentration of fructose 2,6-bisphosphate controlled to rise and fall with blood-glucose levels? Two enzymes regulate the concentration of this molecule: one phosphorylates fructose 6-phosphate and the other dephosphorylates fructose 2,6-bisphosphate. Fructose 2,6-bisphosphate is formed in a reaction catalyzed by *phosphofructokinase 2* (PFK2), a different

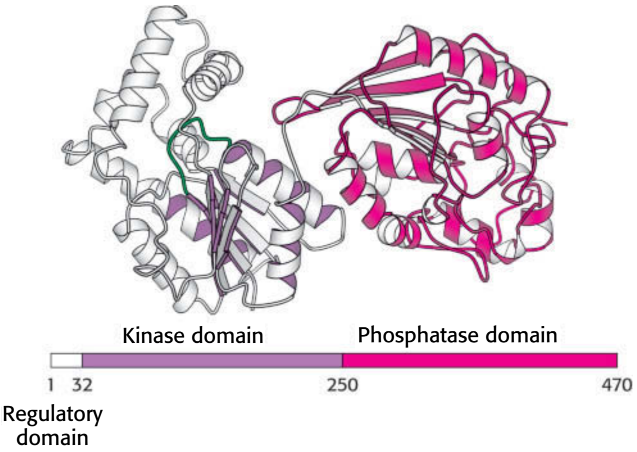

Kinase domain
Phosphatase domain
1
32
250
470
Regulatory domain

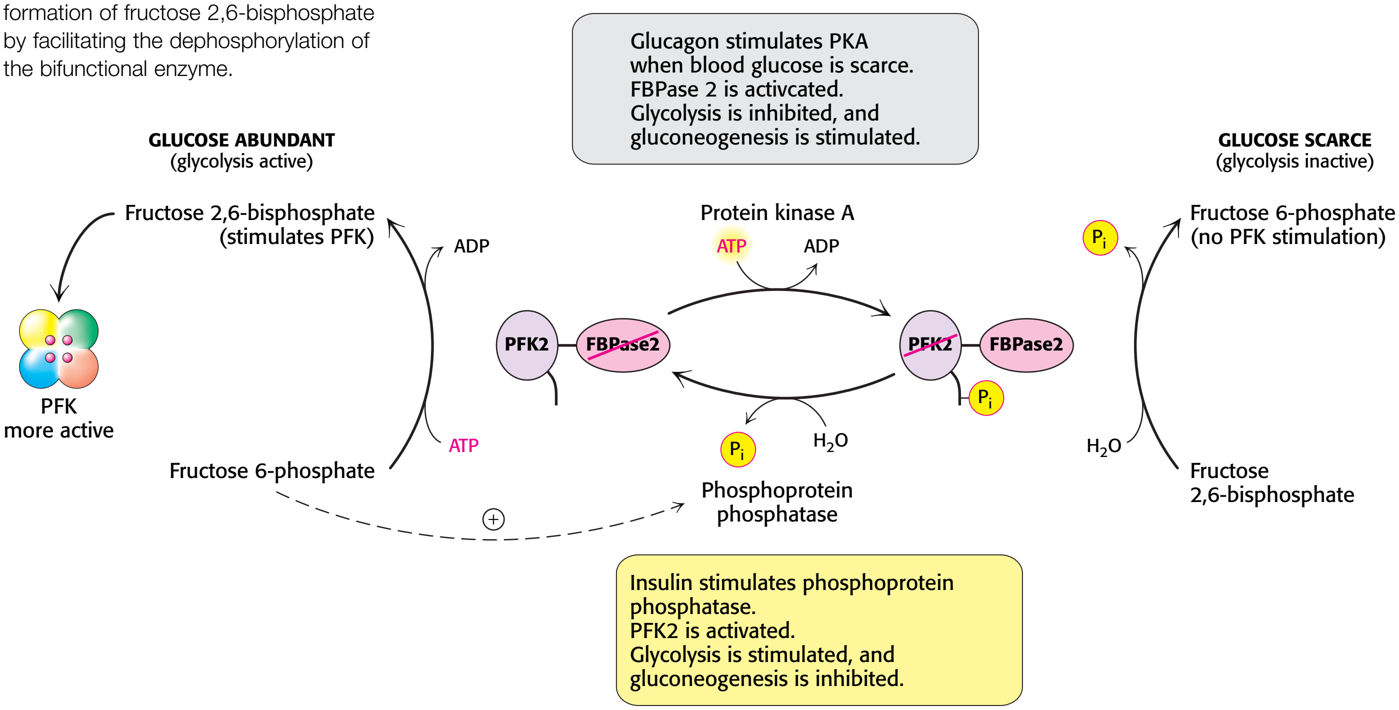
FIGURE 16.30 Domain structure of the bifunctional enzyme phosphofructokinase 2. The kinase domain (purple) is fused to the phosphatase domain (red). The kinase domain is a P-loop NTP hydrolase domain, as indicated by the purple shading (Section 9.4). The bar represents the amino acid sequence of the enzyme. [Drawn from 1BIF.pdb.]

enzyme from phosphofructokinase. Fructose 6-phosphate is formed through the hydrolysis of fructose 2,6-bisphosphate by a specific phosphatase, *fructose bisphosphatase 2* (FBPase2). The striking finding is that *both PFK2 and FBPase2 are present in a single 55-kDa polypeptide chain* (Figure 16.30). This *bifunctional enzyme* contains an N-terminal *regulatory domain,* followed by a *kinase domain* and a *phosphatase domain.* PFK2 resembles adenylate kinase in having a P-loop NTPase domain (Section 9.4), whereas FBPase2 resembles phosphoglycerate mutase (p. 460). Recall that the mutase is essentially a phosphatase. In the bifunctional enzyme, the phosphatase activity evolved to become specific for F-2,6-BP. The bifunctional enzyme itself probably arose by the fusion of genes encoding the kinase and phosphatase domains.

What controls whether PFK2 or FBPase2 dominates the bifunctional enzyme's activities in the liver? The activities of PFK2 and FBPase2 are reciprocally controlled by *phosphorylation of a single serine residue.* When glucose is scarce, such as during a night's fast, a rise in the blood level of the hormone glucagon triggers a cyclic AMP signal cascade (Section 14.1), leading to the phosphorylation of this bifunctional enzyme by protein kinase A (Figure 16.31). This covalent modification activates FBPase2 and inhibits PFK2, lowering the level of F-2,6-BP. Gluconeogenesis predominates. Glucose formed by the liver under these conditions is essential for the viability of the brain. Glucagon stimulation of protein kinase A also inactivates pyruvate kinase in the liver (p. 473).

Conversely, when blood-glucose levels are high, such as after a meal, gluconeogenesis is not needed. Insulin is secreted and initiates a signal pathway that activates a protein phosphatase, which removes the phosphoryl group from the bifunctional enzyme. This covalent modification activates PFK2 and inhibits FBPase2. The resulting rise in the level of F-2,6-BP accelerates glycolysis. In liver, a key role of glycolysis is to generate metabolites for

FIGURE 16.31 Control of the synthesis and degradation of fructose 2,6-bisphosphate. A low blood-glucose level as signaled by glucagon leads to the phosphorylation of the bifunctional enzyme and hence to a lower level of fructose 2,6-bisphosphate, slowing glycolysis. Insulin accelerates the formation of fructose 2,6-bisphosphate by facilitating the dephosphorylation of the bifunctional enzyme.


Glucagon stimulates PKA
when blood glucose is scarce.
FBPase 2 is activcated.
Glycolysis is inhibited, and
gluconeogenesis is stimulated.
GLUCOSE ABUNDANT
(glycolysis active)
GLUCOSE SCARCE
(glycolysis inactive)
Fructose 2,6-bisphosphate
(stimulates PFK)
ADP
Protein kinase A
ATP
ADP
P_i
Fructose 6-phosphate
(no PFK stimulation)
PFK
more active
PFK2
FBPase2
PFK2
FBPase2
P_i
ATP
P_i
H_2O
H_2O
Fructose 6-phosphate
Fructose
2,6-bisphosphate
+
Phosphoprotein
phosphatase
Insulin stimulates phosphoprotein
phosphatase.
PFK2 is activated.
Glycolysis is stimulated, and
gluconeogenesis is inhibited.

biosynthesis. The coordinated control of glycolysis and gluconeogenesis is facilitated by the location of the kinase and phosphatase domains on the same polypeptide chain as the regulatory domain.

The hormones insulin and glucagon also regulate the amounts of essential enzymes. These hormones alter gene expression primarily by changing the rate of transcription. Insulin levels rise subsequent to eating, when there is plenty of glucose for glycolysis. To encourage glycolysis, insulin stimulates the expression of phosphofructokinase, pyruvate kinase, and the bifunctional enzyme that makes and degrades F-2,6-BP. Glucagon rises during fasting, when gluconeogenesis is needed to replace scarce glucose. To encourage gluconeogenesis, glucagon inhibits the expression of the three regulated glycolytic enzymes and stimulates instead the production of two key gluconeogenic enzymes, phosphoenolpyruvate carboxykinase and fructose 1,6-bisphosphatase. Transcriptional control in eukaryotes is much slower than allosteric control, taking hours or days instead of seconds to minutes.

Substrate cycles amplify metabolic signals and produce heat

A pair of reactions such as the phosphorylation of fructose 6-phosphate to fructose 1,6-bisphosphate and its hydrolysis back to fructose 6-phosphate is called a *substrate cycle.* As already mentioned, both reactions are not fully active at the same time because of reciprocal allosteric controls. However, isotope-labeling studies have shown that some fructose 6-phosphate is phosphorylated to fructose 1,6-bisphosphate even during gluconeogenesis. There also is a limited degree of cycling in other pairs of opposed irreversible reactions. This cycling was regarded as an imperfection in metabolic control, and so substrate cycles have sometimes been called *futile cycles.* In malignant hyperthermia, there is rapid uncontrolled hydrolysis of ATP, which generates heat and can raise the body temperature to 44 °C (111°F). Muscles may become rigid and suffer substantial tissue damage.

Despite such extraordinary circumstances, substrate cycles now seem likely to be biologically important. One possibility is that *substrate cycles amplify metabolic signals.* Suppose that the rate of conversion of A into B is 100 and of B into A is 90, giving an initial net flux of 10. Assume that an allosteric effector increases the A → B rate by 20% to 120 and reciprocally decreases the B → A rate by 20% to 72. The new net flux is 48, and so a 20% change in the rates of the opposing reactions has led to a 380% increase in the net flux. In the example shown in Figure 16.32, this amplification is made possible by the rapid hydrolysis of ATP. The flux down the glycolytic pathway has been suggested to increase as much as 1000-fold at the initiation of intense exercise. Because the allosteric activation of enzymes alone seems unlikely to explain this increased flux, the existence of substrate cycles may partly account for the rapid rise in the rate of glycolysis.

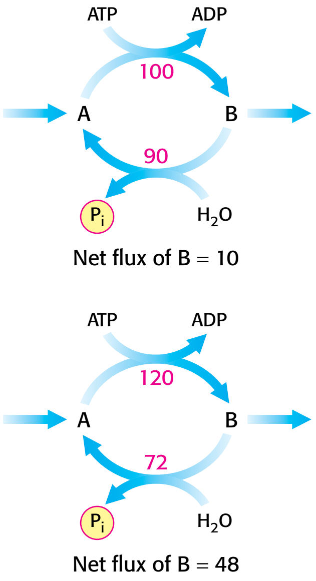


FIGURE 16.32 Substrate cycle. This ATP-driven cycle operates at two different rates. A small change in the rates of the two opposing reactions results in a large change in the *net* flux of product B.

Lactate and alanine formed by contracting muscle are used by other organs

Lactate produced by active skeletal muscle and erythrocytes is a source of energy for other organs. Erythrocytes lack mitochondria and can never oxidize glucose completely. In contracting fast-twitch skeletal muscle fibers during vigorous exercise, the rate at which glycolysis produces pyruvate exceeds the rate at which the citric acid cycle oxidizes it. In these cells, lactate dehydrogenase reduces excess pyruvate to lactate to restore redox balance (p. 464). However, lactate is a dead end in metabolism. It must be converted back into pyruvate before it can be metabolized. Both pyruvate and lactate diffuse out of these cells through carriers into the

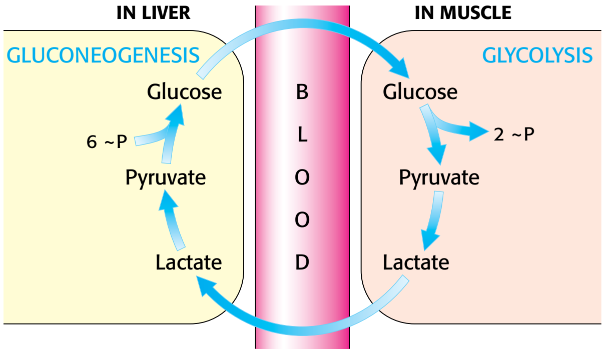
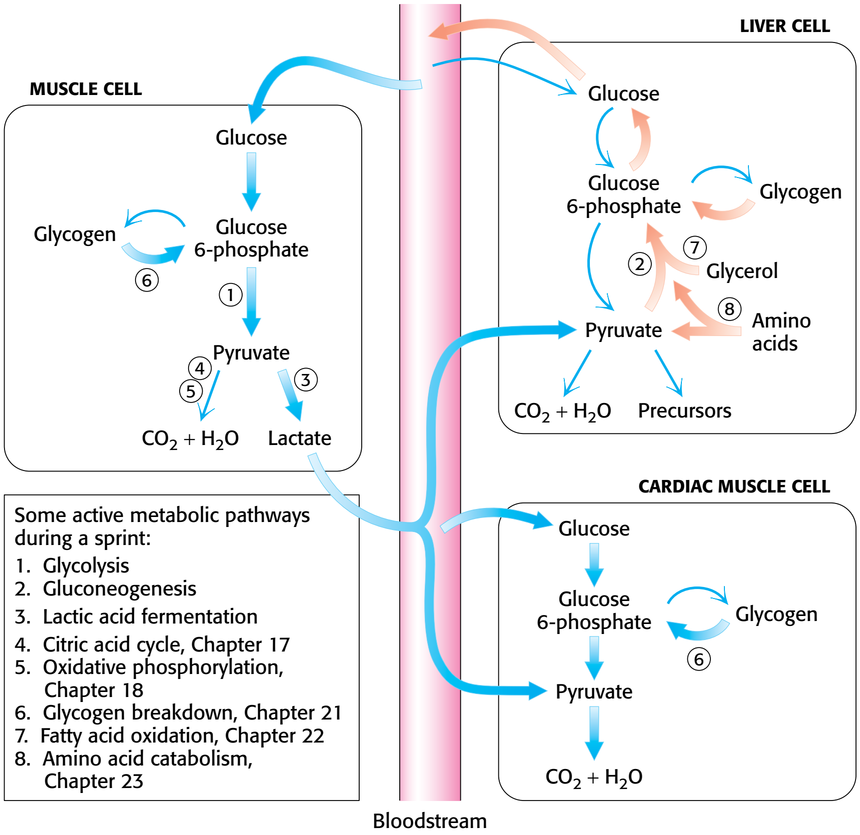

IN LIVER
IN MUSCLE
GLUCONEOGENESIS
GLYCOLYSIS
Glucose
6 ~P
Pyruvate
Lactate
B
L
O
O
D
Glucose
2 ~P
Pyruvate
Lactate

FIGURE 16.33 The Cori cycle. Lactate formed by active muscle is converted into glucose by the liver. This cycle shifts part of the metabolic burden of active muscle to the liver. The symbol ~P represents nucleoside triphosphates.

blood. *In contracting skeletal muscle, the formation and release of lactate lets the muscle generate ATP in the absence of oxygen and shifts the burden of metabolizing lactate from muscle to other organs.* The pyruvate and lactate in the bloodstream have two fates. In one fate, the plasma membranes of some cells, particularly cells in cardiac muscle and slow-twitch (type 1) skeletal muscle, contain carriers that make the cells highly permeable to lactate and pyruvate. These molecules diffuse from the blood into such permeable cells. Once inside these well-oxygenated cells, lactate can be reverted back to pyruvate and metabolized through the citric acid cycle and oxidative phosphorylation to generate ATP. The use of lactate in place of glucose by these cells makes more circulating glucose available to the active muscle cells. In the other fate, excess lactate enters the liver and is converted first into pyruvate and then into glucose by the gluconeogenic pathway. *Contracting skeletal muscle supplies lactate to the liver, which uses it to synthesize and release glucose. Thus, the liver restores the level of glucose necessary for active muscle cells, which derive ATP from the glycolytic conversion of glucose into lactate. These reactions constitute the Cori cycle* (Figure 16.33).

Studies have shown that alanine, like lactate, is a major precursor of glucose in the liver. The alanine is generated in muscle when the carbon skeletons of some amino acids are used as fuels. The nitrogens from these amino acids are transferred to pyruvate to form alanine; the reverse reaction takes place in the liver. This process also helps maintain nitrogen balance. The interplay between glycolysis and gluconeogenesis is summarized in Figure 16.34, which shows how these pathways help meet the energy needs of different cell types.

FIGURE 16.34 PATHWAY INTEGRATION: Cooperation between glycolysis and gluconeogenesis during a sprint. Glycolysis and gluconeogenesis are coordinated, in a tissue-specific fashion, to ensure that the energy needs of all cells are met. Consider a sprinter. In skeletal leg muscle, glucose will be metabolized aerobically to CO_2 and H_2O or, more likely (thick arrows) during a sprint, anaerobically to lactate. In cardiac muscle, the lactate can be converted into pyruvate and used as a fuel, along with glucose, to power the heartbeats to keep the sprinter's blood flowing. Gluconeogenesis, a primary function of the liver, will be taking place rapidly (thick arrows) to ensure that enough glucose is present in the blood for skeletal and cardiac muscle, as well as for other tissues. Glycogen, glycerol, and amino acids are other sources of energy that we will learn about in later chapters.


MUSCLE CELL
Glucose
Glycogen
Glucose 6-phosphate
⑥
①
Pyruvate
④
⑤
③
$CO_2 + H_2O$
Lactate
LIVER CELL
Glucose
Glucose 6-phosphate
Glycogen
②
⑦
Glycerol
⑧
Amino acids
Pyruvate
$CO_2 + H_2O$
Precursors
CARDIAC MUSCLE CELL
Glucose
Glucose 6-phosphate
Glycogen
⑥
Pyruvate
$CO_2 + H_2O$
Bloodstream

Some active metabolic pathways during a sprint:

1. Glycolysis
2. Gluconeogenesis
3. Lactic acid fermentation
4. Citric acid cycle, Chapter 17
5. Oxidative phosphorylation, Chapter 18
6. Glycogen breakdown, Chapter 21
7. Fatty acid oxidation, Chapter 22
8. Amino acid catabolism, Chapter 23

Isozymic forms of lactate dehydrogenase in different tissues catalyze the interconversions of pyruvate and lactate (Section 10.2). Lactate dehydrogenase is a tetramer of two kinds of 35-kDa subunits encoded by similar genes: the H type predominates in the heart, and the homologous M type in skeletal muscle and the liver. These subunits associate to form five types of tetramers: H_4, H_3M_1, H_2M_2, H_1M_3, and M_4. The H_4 isozyme (type 1) has higher affinity for substrates than that of the M_4 isozyme (type 5) and, unlike M_4, is allosterically inhibited by high levels of pyruvate. The other isozymes have intermediate properties, depending on the ratio of the two kinds of chains. The H_4 isozyme oxidizes lactate to pyruvate, which is then used as a fuel by the heart through aerobic metabolism. Indeed, heart muscle never functions anaerobically. In contrast, M_4 is optimized to operate in the reverse direction, to convert pyruvate into lactate to allow glycolysis to proceed under anaerobic conditions. Interestingly, endurance training increases the H_4 isozyme in slow-twitch muscle fibers, enabling these fibers to use the lactate generated by other fiber types.

Glycolysis and gluconeogenesis are evolutionarily intertwined

The metabolism of glucose has ancient origins. Organisms living in the early biosphere depended on the anaerobic generation of energy until significant amounts of oxygen began to accumulate 2 billion years ago. Glycolytic enzymes were most likely derived independently rather than by gene duplication, because glycolytic enzymes with similar properties do not have similar amino acid sequences. Although there are four kinases and two isomerases in the pathway, sequence and structural comparisons suggest that these sets of enzymes are not related to one another by divergent evolution. The common dinucleotide-binding domain found in the dehydrogenases (Figure 16.12) and the $\alpha\beta$ barrels are the only major recurring elements.

We can speculate on the relationship between glycolysis and gluconeogenesis if we think of glycolysis as consisting of two segments: the metabolism of hexoses (the upper segment) and the metabolism of trioses (the lower segment). The enzymes of the upper segment are different in some species and are missing entirely in some archaea, whereas enzymes of the lower segment are quite conserved. In fact, four enzymes of the lower segment are present in all species. *This lower part of the pathway is common to glycolysis and gluconeogenesis.* This common part of the two pathways may be the oldest part, constituting the core to which the other steps were added. The upper part would have varied according to the sugars that were available to evolving organisms in particular niches. Interestingly, this core part of carbohydrate metabolism can generate triose precursors for ribose sugars, a component of RNA and a critical requirement for the RNA world. Thus, we are left with the unanswered question, Was the original core pathway used for energy conversion or biosynthesis?

SUMMARY

16.1 Glycolysis Is an Energy-Conversion Pathway in Many Organisms

Glycolysis is the set of reactions that converts glucose into pyruvate. The 10 reactions of glycolysis take place in the cytoplasm. In the first stage, glucose is converted into fructose 1,6-bisphosphate by a phosphorylation, an isomerization, and a second phosphorylation reaction. Fructose 1,6-bisphosphate is then cleaved by aldolase into dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, which are readily interconvertible. Two molecules of ATP are consumed per molecule of glucose in these reactions. In the second stage, ATP is generated. Glyceraldehyde 3-phosphate is oxidized and phosphorylated to form 1,3-bisphosphoglycerate, an acyl phosphate with a high phosphoryl-transfer potential. This molecule transfers a phosphoryl group to ADP to form ATP and 3-phosphoglycerate. A phosphoryl shift and a dehydration form phosphoenolpyruvate, a second intermediate with a high phosphoryl-transfer potential. Another molecule of ATP is generated as phosphoenolpyruvate is converted into pyruvate. There is a net gain of two molecules of ATP in the formation of two molecules of pyruvate from one molecule of glucose.

The electron acceptor in the oxidation of glyceraldehyde 3-phosphate is NAD^+, which must be regenerated for glycolysis to continue. In aerobic organisms, the NADH formed in glycolysis transfers its electrons to O_2 through the electron-transport chain, which thereby regenerates NAD^+. Under anaerobic conditions and in some microorganisms, NAD^+ is regenerated by the reduction of pyruvate to lactate. In other microorganisms, NAD^+ is regenerated by the reduction of pyruvate to ethanol. These two processes are examples of fermentations.

16.2 The Glycolytic Pathway Is Tightly Controlled

The glycolytic pathway has a dual role: it degrades glucose to generate ATP, and it provides building blocks for the synthesis of cellular components. The rate of conversion of glucose into pyruvate is regulated to meet these two major cellular needs. Under physiological conditions, the reactions of glycolysis are readily reversible except for those catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase. Phosphofructokinase, the most important control element in glycolysis, is inhibited by high levels of ATP and citrate, and it is activated by AMP and fructose 2,6-bisphosphate. In the liver, this bisphosphate signals that glucose is abundant. Hence, phosphofructokinase is active when either energy or building blocks are needed. Hexokinase is inhibited by glucose 6-phosphate, which accumulates when phosphofructokinase is inactive. ATP and alanine allosterically inhibit pyruvate kinase, the other control site, and fructose 1,6-bisphosphate activates the enzyme. Consequently, pyruvate kinase is maximally active when the energy charge is low and glycolytic intermediates accumulate.

16.3 Glucose Can Be Synthesized from Noncarbohydrate Precursors

Gluconeogenesis, which occurs primarily in the liver, is the synthesis of glucose from noncarbohydrate sources, such as lactate, amino acids, glycerol, and alanine produced from pyruvate by active skeletal muscle. Several of the reactions that convert pyruvate into glucose are common to glycolysis. Gluconeogenesis, however, requires four new reactions to bypass the essential irreversibility of three reactions in glycolysis. In two of the new reactions, pyruvate is carboxylated in mitochondria to oxaloacetate, which in turn is decarboxylated and phosphorylated in the cytoplasm to phosphoenolpyruvate. ATP and GTP are consumed in these reactions, which are catalyzed by pyruvate carboxylase and phosphoenolpyruvate carboxykinase, respectively. The other distinctive reactions of gluconeogenesis are the hydrolyses of fructose 1,6-bisphosphate and glucose 6-phosphate, which are catalyzed by specific phosphatases.

16.4 Gluconeogenesis and Glycolysis Are Reciprocally Regulated

Gluconeogenesis and glycolysis are reciprocally regulated so that one pathway is relatively inactive while the other is highly active. Phosphofructokinase and fructose 1,6-bisphosphatase are key control points. Fructose 2,6-bisphosphate, an intracellular signal molecule present at higher levels when glucose is abundant, activates glycolysis and inhibits gluconeogenesis by regulating these enzymes. Pyruvate kinase and pyruvate carboxylase are regulated by other effectors so that both are not maximally active at the same time. Allosteric regulation and reversible phosphorylation, which are rapid, are complemented by transcriptional control, which takes place in hours or days.

KEY TERMS

glycolysis (p. 449)
lactic acid fermentation (p. 449)
alcoholic fermentation (p. 449)
gluconeogenesis (p. 449)
hexokinase (p. 451)
kinase (p. 453)
phosphofructokinase (PFK) (p. 454)
thioester intermediate (p. 457)
substrate-level phosphorylation (p. 460)
pyruvate kinase (p. 461)
fermentation (p. 462)
obligate anaerobe (p. 464)
facultative anaerobe (p. 464)
Rossmann fold (p. 465)
committed step (p. 471)
feedforward stimulation (p.472)
glucokinase (p. 473)
aerobic glycolysis (p. 474)
pyruvate carboxylase (p. 478)
biotin (p. 479)
phosphoenolpyruvate carboxykinase (p. 480)
fructose 1,6-bisphosphatase (p. 481)
glucose 6-phosphatase (p. 481)
bifunctional enzyme (p. 484)
substrate cycle (p. 485)
Cori cycle (p. 486)

PROBLEMS

1. *Gross versus net*. The gross yield of ATP from the metabolism of glucose to two molecules of pyruvate is four molecules of ATP. However, the net yield is only two molecules of ATP. Why are the gross and net values different?

2. *Go together like the owl and the pussycat*. Match each term with its description.

(a) Hexokinase ________	1. Forms fructose 1,6-bisphosphate
(b) Phosphoglucose isomerase ________	2. Generates the first high-phosphoryl-transfer-potential compound that is not ATP
(c) Phosphofructokinase ________	3. Converts glucose 6-phosphate into fructose 6-phosphate
(d) Aldolase ________	4. Phosphorylates glucose
(e) Triose phosphate isomerase ________	5. Generates the second molecule of ATP
(f) Glyceraldehyde 3-phosphate dehydrogenase ________	6. Cleaves fructose 1, 6-bisphosphate
(g) Phosphoglycerate kinase ________	7. Generates the second high-phosphoryl-transfer-potential compound that is not ATP
(h) Phosphoglycerate mutase ________	8. Catalyzes the interconversion of three-carbon isomers
(i) Enolase ________	9. Converts 3-phosphoglycerate into 2-phosphoglycerate
(j) Pyruvate kinase ________	10. Generates the first molecule of ATP

3. *Who takes? Who gives?* Lactic acid fermentation and alcoholic fermentation are oxidation–reduction reactions. Identify the ultimate electron donor and electron acceptor.

4. *ATP yield.* Each of the following molecules is processed by glycolysis to lactate. How much ATP is generated from each molecule?

(a) Glucose 6-phosphate
(b) Dihydroxyacetone phosphate
(c) Glyceraldehyde 3-phosphate
(d) Fructose
(e) Sucrose

5. *Enzyme redundancy?* Why is it advantageous for the liver to have both hexokinase and glucokinase to phosphorylate glucose?

6. *Magic?* The interconverison of DHAP and GAP greatly favors the formation of DHAP at equilibirum. Yet the conversion of DHAP by triose phosphate isomerase proceeds readily. Why?

7. *Between two extremes*. What is the role of a thioester in the formation of ATP in glycolysis?

8. *Corporate sponsors*. Some of the early research on glycolysis was supported by the brewing industry. Why would the brewing industry be interested in glycolysis?

9. *Recommended daily allowance*. The recommended daily allowance for the vitamin niacin is 15 mg per day. How would glycolysis be affected by niacin deficiency?

10. *Who's on first?* Although both hexokinase and phosphofructokinase catalyze irreversible steps in glycolysis and the hexokinase-catalyzed step is first, phosphofructokinase

is nonetheless the pacemaker of glycolysis. What does this information tell you about the fate of the glucose 6-phosphate formed by hexokinase?

11. *The tortoise and the hare.* Why is the regulation of phosphofructokinase by energy charge not as important in the liver as it is in muscle?

12. *Like Minneapolis and St. Paul.* Match each term with its description.

(a) Lactate ______
(b) Pyruvate carboxylase ______
(c) Acetyl CoA ______
(d) Phosphoenolpyruvate carboxykinase ______
(e) Glycerol ______
(f) Fructose 1,6-bisphosphatase ______
(g) Glucose 6-phosphatase ______

1. Generates oxaloacetate
2. Readily converted into DHAP
3. Generates a high-phosphoryl-transfer-potential compound
4. Found predominantly in liver
5. Gluconeogenic counterpart of PFK
6. Readily converted into pyruvate
7. Required for pyruvate carboxylase activity

13. *Running in reverse.* Why can't the reactions of the glycolytic pathway simply be run in reverse to synthesize glucose?

14. *Road blocks.* What reactions of glycolysis are not readily reversible under intracellular conditions?

15. *No pickling.* Why is it in the muscle's best interest to export lactic acid into the blood during intense exercise?

16. *Après vous.* Why is it physiologically advantageous for the pancreas to use GLUT2, with a high K_M, as the transporter that allows glucose entry into β cells?

17. *Bypass.* In the liver, fructose can be converted into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate without passing through the phosphofructokinase-regulated reaction. Show the reactions that make this conversion possible. Why might ingesting high levels of fructose have deleterious physiological effects?

18. *Match 'em 1.* The following sequence is a part of the sequence of reactions in gluconeogenesis.

$$\text{Pyruvate} \xrightarrow[\text{A}]{} \text{Oxaloacetate} \xrightarrow[\text{B}]{} \text{Malate} \xrightarrow[\text{C}]{} \text{Oxaloacetate} \xrightarrow[\text{D}]{} \text{Phosphoenolpyruvate}$$

Match the capital letters representing the reaction in the gluconeogenic pathway with parts *a, b, c,* etc.

(a) takes place in mitochondria
(b) takes place in the cytoplasm
(c) produces CO_2
(d) consumes CO_2
(e) requires NADH
(f) produces NADH
(g) requires ATP
(h) requires GTP
(i) requires thiamine
(j) requires biotin
(k) is regulated by acetyl CoA.

19. *Like Batman and Robin.* Match parts *a* through *j* with parts 1 through 10.

(a) Glucose 6-phosphate ______
(b) [ATP] < [AMP] ______
(c) Citrate ______
(d) Low pH ______
(e) Fructose 1, 6-bisphosphate ______
(f) Fructose 2, 6 bisphosphate ______
(g) Insulin ______
(h) Has a high K_M for glucose ______
(i) Transporter specific to liver and pancreas ______
(j) [ATP] > [AMP] ______

1. Inhibits phosphofructokinase in the liver
2. Glucokinase
3. GLUT2
4. Inhibits hexokinase
5. Inhibits phosphofructokinase in muscle
6. Inhibits phosphofructokinase
7. Stimulates pyruvate kinase
8. Stimulates phosphofructokinase in the liver
9. Causes the insertion of GLUT4 into cell membranes
10. Stimulates phosphofructokinase

20. *Trouble ahead.* Suppose that a microorganism that was an obligate anaerobe suffered a mutation that resulted in the loss of triose phosphate isomerase activity. How would this loss affect the ATP yield of fermentation? Could such an organism survive?

21. *Kitchen chemistry.* Sucrose is commonly used to preserve fruits. Why is glucose not suitable for preserving foods?

22. *Tracing carbon atoms 1.* Glucose labeled with ^{14}C at C-1 is incubated with the glycolytic enzymes and necessary cofactors.

(a) What is the distribution of ^{14}C in the pyruvate that is formed? (Assume that the interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate is very rapid compared with the subsequent step.)

(b) If the specific activity of the glucose substrate is 10 mCi $mmol^{-1}$ (millicuries per mole, a measure of radioactivity per mole), what is the specific activity of the pyruvate that is formed?

23. *Lactic acid fermentation.*

(a) Write a balanced equation for the conversion of glucose into lactate.

(b) Calculate the standard free-energy change of this reaction by using the data given in Table 16.1 and the fact that $\Delta G^{\circ\prime}$ is -25 kJ mol^{-1} (-6 kcal mol^{-1}) for the following reaction:

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{lactate} + \text{NAD}^+$$

What is the free-energy change (ΔG, not $\Delta G^{\circ\prime}$) of this reaction when the concentrations of reactants are: glucose, 5 mM; lactate, 0.05 mM; ATP, 2 mM; ADP, 0.2 mM; and P_i, 1 mM?

24. *High potential.* What is the equilibrium ratio of phosphoenolpyruvate to pyruvate under standard conditions when [ATP]/[ADP] = 10?

25. *Hexose–triose equilibrium.* What are the equilibrium concentrations of fructose 1,6-bisphosphate, dihydroxyacetone phosphate, and glyceraldehyde 3-phosphate when 1 mM fructose 1,6-bisphosphate is incubated with aldolase under standard conditions?

26. *Double labeling.* 3-Phosphoglycerate labeled uniformly with ^{14}C is incubated with 1,3-BPG labeled with ^{32}P at C-1. What is the radioisotope distribution of the 2,3-BPG that is formed on addition of BPG mutase?

27. *An informative analog.* Xylose has the same structure as that of glucose except that it has a hydrogen atom at G-5 in place of a hydroxymethyl group. The rate of ATP hydrolysis by hexokinase is markedly enhanced by the addition of xylose. Why?

28. *Distinctive sugars.* The intravenous infusion of fructose into healthy volunteers leads to a two- to fivefold increase in the level of lactate in the blood, a far greater increase than that observed after the infusion of the same amount of glucose.

(a) Why is glycolysis more rapid after the infusion of fructose?

(b) Fructose has been used in place of glucose for intravenous feeding. Why is this use of fructose unwise?

29. *It is not hard to meet expenses. They are everywhere.* What energetic barrier prevents glycolysis from simply running in reverse to synthesis glucose? What is the energetic cost to overcome this barrier?

30. *Match 'em 2.* Indicate which of the conditions listed in the right-hand column increase the activity of the glycolytic and gluconeogenic pathways.

(a) glycolysis ________
(b) gluconeogenesis ________

1. increase in ATP
2. increase in AMP
3. increase in fructose 2,6-bisphosphate
4. increase in citrate
5. increase in acetyl CoA
6. increase in insulin
7. increase in glucagon
8. fasting
9. fed

31. *Waste not, want not.* Why is the conversion of lactic acid from the blood into glucose in the liver in an organism's best interest?

32. *Road blocks bypassed.* How are the irreversible reactions of glycolysis bypassed in gluconeogenesis?

33. *Pointlessness averted.* What are the regulatory means that prevent high levels of activity in glycolysis and gluconeogenesis simultaneously?

34. *Different needs.* Liver is primarily a gluconeogenic tissue, whereas muscle is primarily glycolytic. Why does this division of labor make good physiological sense?

35. *Metabolic mutants.* What would be the effect on an organism's ability to use glucose as an energy source if a mutation inactivated glucose 6-phosphatase in the liver?

36. *Never let me go.* Why does the lack of glucose 6-phosphatase activity in the brain and muscle make good physiological sense?

37. *Counting high-energy compounds 1.* How many NTP molecules are required for the synthesis of one molecule of glucose from two molecules of pyruvate? How many NADH molecules?

38. *Counting high-energy compounds 2.* How many NTP molecules are required to synthesize glucose from each of the following compounds?

(a) Glucose 6-phosphate

(b) Fructose 1,6-bisphosphate

(c) Two molecules of oxaloacetate

(d) Two molecules of dihydroxyacetone phosphate

39. *Lending a hand.* How might enzymes that remove amino groups from alanine and aspartate contribute to gluconeogenesis?

40. *More metabolic mutants.* Predict the effect of each of the following mutations on the pace of glycolysis in liver cells:

(a) Loss of the allosteric site for ATP in phosphofructokinase

(b) Loss of the binding site for citrate in phosphofructokinase

(c) Loss of the phosphatase domain of the bifunctional enzyme that controls the level of fructose 2,6-bisphosphate

(d) Loss of the binding site for fructose 1,6-bisphosphate in pyruvate kinase

41. *Yet another metabolic mutant.* What are the likely consequences of a genetic disorder rendering fructose 1,6-bisphosphatase in the liver less sensitive to regulation by fructose 2,6-bisphosphate?

42. *Biotin snatcher.* Avidin, a 70-kDa protein in egg white, has very high affinity for biotin. In fact, it is a highly specific inhibitor of biotin enzymes. Which of the following

conversions would be blocked by the addition of avidin to a cell homogenate?

(a) Glucose → pyruvate
(b) Pyruvate → glucose
(c) Oxaloacetate → glucose
(d) Malate → oxaloacetate
(e) Pyruvate → oxaloacetate
(f) Glyceraldehyde 3-phosphate → fructose 1,6-bisphosphate

43. *Tracing carbon atoms 2.* If cells synthesizing glucose from lactate are exposed to CO_2 labeled with ^{14}C, what will be the distribution of label in the newly synthesized glucose?

44. *Arsenate poisoning.* Arsenate.(AsO_4^{3-}) closely resembles P_i in structure and reactivity. In the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase, arsenate can replace phosphate in attacking the energy-rich thioester intermediate. The product of this reaction, 1-arseno-3-phosphoglycerate, is unstable. It and other acyl arsenates are rapidly and spontaneously hydrolyzed. What is the effect of arsenate on energy generation in a cell?

45. *Reduce, reuse, recycle.* In the conversion of glucose into two molecules of lactate, the NADH generated earlier in the pathway is oxidized to NAD^+. Why is it not to the cell's advantage to simply make more NAD^+ so that the regeneration would not be necessary? After all, the cell would save much energy because it would no longer need to synthesize lactic acid dehydrogenase.

46. *Adenylate kinase again.* Adenylate kinase, an enzyme considered in great detail in Chapter 9, is responsible for interconverting the adenylate nucleotide pool:

$$\text{ADP} + \text{ADP} \rightleftharpoons \text{ATP} + \text{AMP}$$

The equilibrium constant for this reaction is close to 1, inasmuch as the number of phosphoanhydride bonds is the same on each side of the equation. Using the equation for the equilibrium constant for this reaction, show why changes in [AMP] are a more effective indicator of the adenylate pool than [ATP].

47. *Working at cross-purposes?* Gluconeogenesis takes place during intense exercise, which seems counterintuitive. Why would an organism synthesize glucose and at the same time use glucose to generate energy?

48. *Powering pathways.* Compare the stoichiometries of glycolysis and gluconeogenesis. Recall that the input of one ATP equivalent changes the equilibrium constant of a reaction by a factor of about 10^8 (Section 15.2). By what factor do the additional high-phosphoryl-transfer compounds alter the equilibrium constant of gluconeogenesis?

Mechanism Problem

49. *Argument by analogy.* Propose a mechanism for the conversion of glucose 6-phosphate into fructose 6-phosphate by phosphoglucose isomerase based on the mechanism of triose phosphate isomerase.

Chapter Integration Problems

50. *Not just for energy.* People with galactosemia display central nervous system abnormalities even if galactose is eliminated from the diet. The precise reason for it is not known. Suggest a plausible explanation.

51. *Power to the rbc!* Hexokinase in red blood cells has a K_M of approximately 50 μM. Because life is hard enough as it is, let's assume that the hexokinase displays Michaelis–Menten kinetics. What concentration of blood glucose would yield v_o equal to 90% V_{max}? What does this result tell you if normal blood glucose levels range between approximately 3.6 and 6.1 mM?

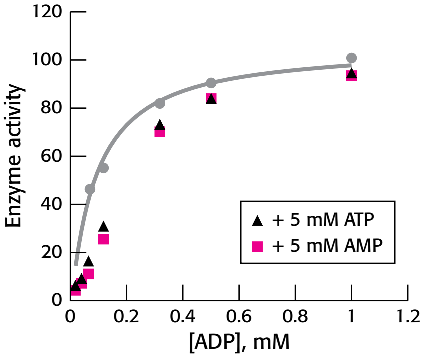
52. *State function.* Fructose 2,6-bisphosphate is a potent stimulator of phosphofructokinase. Explain how fructose 2,6-bisphosphate might function in the concerted model for allosteric enzymes.

Data Interpretation Problems

53. *Now, that's unusual.* Phosphofructokinase has recently been isolated from the hyperthermophilic archaeon *Pyrococcus furiosus.* It was subjected to standard biochemical analysis to determine basic catalytic parameters. The processes under study were of the form

$$\text{Fructose 6-phosphate} + (x - P_i) \rightarrow \text{fructose 1,6-bisphosphate} + (x)$$

The assay measured the increase in fructose 1,6-bisphosphate. Selected results are shown in the adjoining graph.



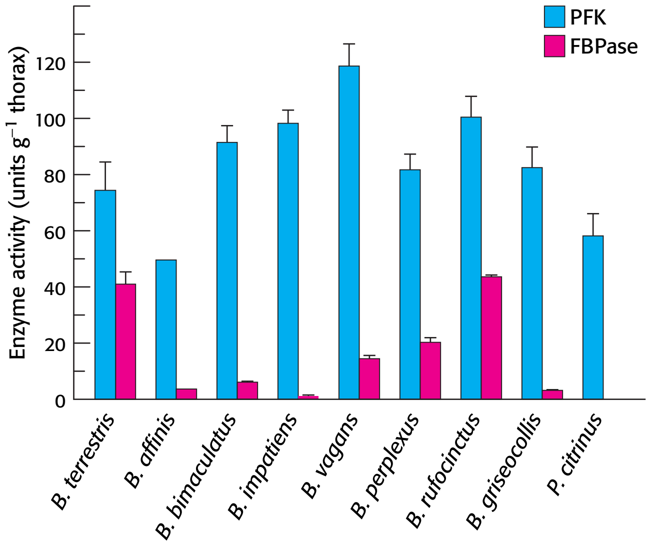
[Data from J. E. Tuininga et al. *J. Biol. Chem.* 274:21023–21028, 1999.]

(a) How does the *P. furiosus* phosphofructokinase differ from the phosphofructokinase considered in this chapter?
(b) What effects do AMP and ATP have on the reaction with ADP?

54. *Cool bees.* In principle, a futile cycle that includes phosphofructokinase and fructose 2,6-bisphosphatase could be used to generate heat. The heat could be used to warm

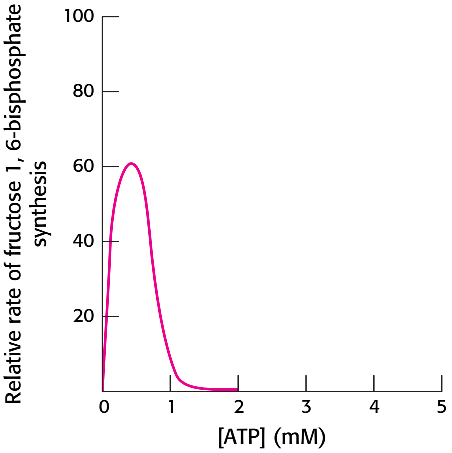
tissues. For instance, certain bumblebees have been reported to use such a futile cycle to warm their flight muscles on cool mornings. Scientists undertook a series of experiments to determine if a number of species of bumblebee use this futile cycle. Their approach was to measure the activity of PFK and F-1,6-BPase in flight muscle.

(a) What was the rationale for comparing the activities of these two enzymes?
(b) The data below show the activites of both enzymes for a variety of bumblebee species (genera *Bombus* and *Psithyrus*). Do these results support the notion that bumblebees use futile cycles to generate heat? Explain.
(c) In which species might futile cycling take place? Explain your reasoning.
(d) Do these results prove that futile cycling does not participate in heat generation?



[Data from J. F. Staples, E. L. Koen, and T. M. Laverty, *J. Exp. Biol.* 207:749–754, 2004, p. 751.]

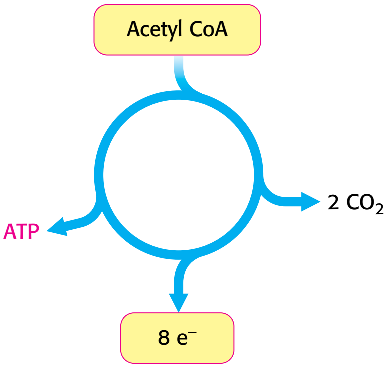
55. *Confused?* The adjoining graph shows muscle phosphofructokinase activity as a function of ATP concentration in the presence of a constant concentration of fructose 6-phosphate. Explain these results and discuss how they relate to the role of phosphofructokinase in glycolysis.



CHAPTER

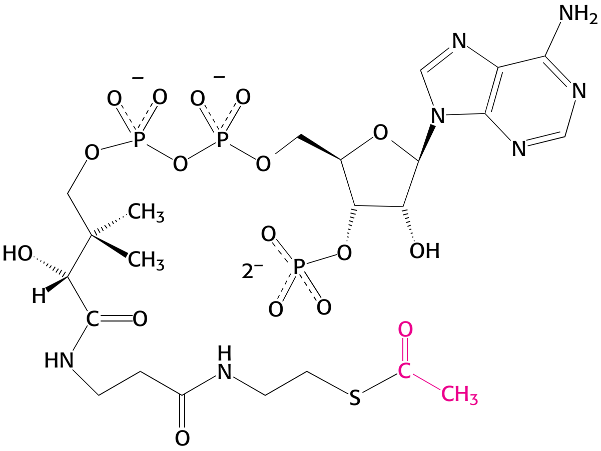
17

The Citric Acid Cycle



Roundabouts, or traffic circles, function as hubs to facilitate traffic flow. The citric acid cycle is the biochemical hub of the cell, oxidizing carbon fuels, usually in the form of acetyl CoA, as well as serving as a source of precursors for biosynthesis. [(Left) Lynn Saville/Getty Images.]

The metabolism of glucose to pyruvate in glycolysis, an anaerobic process, harvests but a fraction of the ATP available from glucose. Most of the ATP generated in metabolism is provided by the *aerobic* processing of glucose. This process starts with the complete oxidation of glucose derivatives to carbon dioxide. This oxidation takes place in a series of reactions called the *citric acid cycle,* also known as the *tricarboxylic acid* (TCA) *cycle* or the *Krebs cycle*. The citric acid cycle is the *final common pathway for the oxidation of fuel molecules*—carbohydrates, fatty acids, and amino acids. Most fuel molecules enter the cycle as *acetyl coenzyme A.*



Acetyl coenzyme A (Acetyl CoA)

OUTLINE

17.1 The Pyruvate Dehydrogenase Complex Links Glycolysis to the Citric Acid Cycle

17.2 The Citric Acid Cycle Oxidizes Two-Carbon Units

17.3 Entry to the Citric Acid Cycle and Metabolism Through It Are Controlled

17.4 The Citric Acid Cycle Is a Source of Biosynthetic Precursors

17.5 The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate

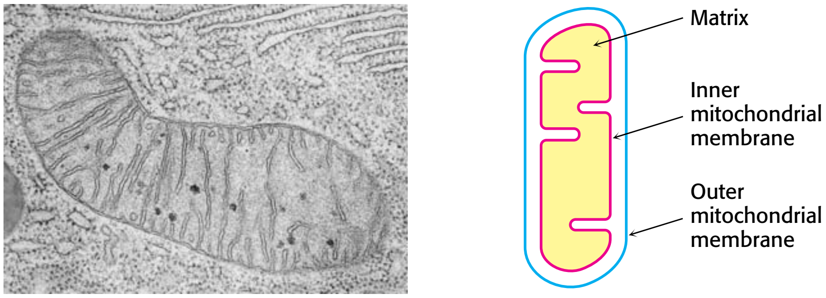

Matrix
Inner mitochondrial membrane
Outer mitochondrial membrane

FIGURE 17.1 Mitochondrion. The double membrane of the mitochondrion is evident in this electron micrograph. The numerous invaginations of the inner mitochondrial membrane are called cristae. The oxidative decarboxylation of pyruvate and the sequence of reactions in the citric acid cycle take place within the matrix. [(Left) Omikron/Photo Researchers.]

Under aerobic conditions, the pyruvate generated from glucose is oxidatively decarboxylated to form acetyl CoA. In eukaryotes, the reactions of the citric acid cycle take place in the matrix of the mitochondria (Figure 17.1), in contrast with those of glycolysis, which take place in the cytoplasm.

The citric acid cycle harvests high-energy electrons

The citric acid cycle is the central metabolic hub of the cell. It is the gateway to the aerobic metabolism of any molecule that can be transformed into an acetyl group or a component of the citric acid cycle. The cycle is also an important source of precursors for the building blocks of many other molecules such as amino acids, nucleotide bases, and porphyrin (the organic component of heme). The citric acid cycle component, oxaloacetate, is also an important precursor to glucose (Section 16.3).

What is the function of the citric acid cycle in transforming fuel molecules into ATP? Recall that fuel molecules are carbon compounds that are capable of being oxidized—that is, of losing electrons (Chapter 15). The citric acid cycle includes a series of oxidation–reduction reactions that result in the oxidation of an acetyl group to two molecules of carbon dioxide. This oxidation generates high-energy electrons that will be used to power the synthesis of ATP. *The function of the citric acid cycle is the harvesting of high-energy electrons from carbon fuels.*

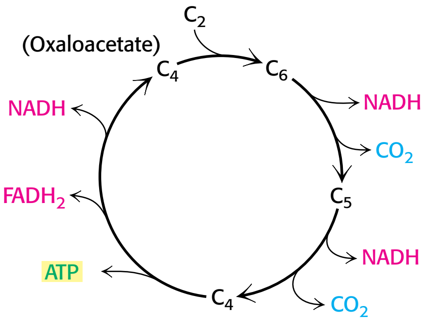

C_2
(Oxaloacetate)
C_4
C_6
NADH
NADH
CO_2
$FADH_2$
C_5
NADH
ATP
C_4
CO_2

FIGURE 17.2 Overview of the citric acid cycle. The citric acid cycle oxidizes two-carbon units, producing two molecules of CO_2, one molecule of ATP, and high-energy electrons in the form of NADH and $FADH_2$.

The overall pattern of the citric acid cycle is shown in Figure 17.2. A four-carbon compound (oxaloacetate) condenses with a two-carbon acetyl unit to yield a six-carbon tricarboxylic acid. The six-carbon compound releases CO_2 twice in two successive oxidative decarboxylations that yield high-energy electrons. A four-carbon compound remains. This four-carbon compound is further processed to regenerate oxaloacetate, which can initiate another round of the cycle. Two carbon atoms enter the cycle as an acetyl unit and two carbon atoms leave the cycle in the form of two molecules of CO_2.

Note that the citric acid cycle itself neither generates a large amount of ATP nor includes oxygen as a reactant (Figure 17.3). Instead, the citric acid

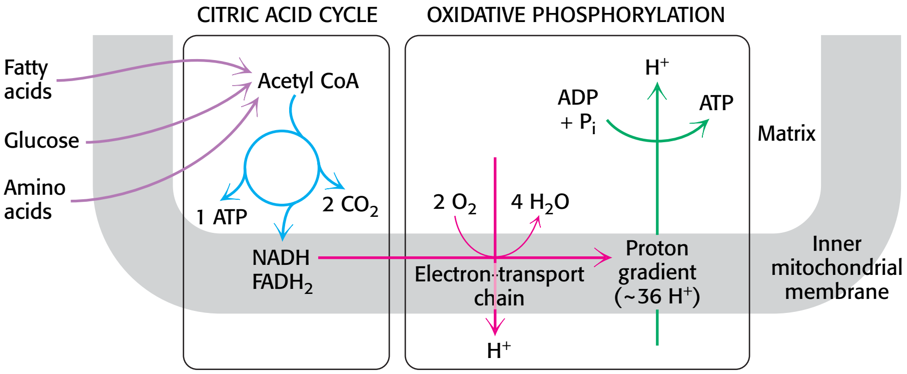

CITRIC ACID CYCLE
OXIDATIVE PHOSPHORYLATION
Fatty acids
Glucose
Amino acids
Acetyl CoA
1 ATP
2 CO_2
NADH
$FADH_2$
H^+
ADP + P_i
ATP
Matrix
2 O_2
4 H_2O
Electron-transport chain
Proton gradient (~36 H^+)
H^+
Inner mitochondrial membrane

FIGURE 17.3 Cellular respiration. The citric acid cycle constitutes the first stage in cellular respiration, the removal of high-energy electrons from carbon fuels in the form of NADH and $FADH_2$ (blue pathway). These electrons reduce O_2 to generate a proton gradient (red pathway), which is used to synthesize ATP (green pathway). The reduction of O_2 and the synthesis of ATP constitute oxidative phosphorylation.

cycle removes electrons from acetyl CoA and uses these electrons to reduce NAD^+ and FAD to form NADH and $FADH_2$. Three hydride ions (hence, six electrons) are transferred to three molecules of nicotinamide adenine dinucleotide (NAD^+), and one pair of hydrogen atoms (hence, two electrons) is transferred to one molecule of flavin adenine dinucleotide (FAD). Electrons released in the reoxidation of NADH and $FADH_2$ flow through a series of membrane proteins (referred to as the *electron-transport chain*) to generate a proton gradient across the inner mitochondrial membrane. These protons then flow through ATP synthase to generate ATP from ADP and inorganic phosphate. These electron carriers yield nine molecules of ATP when they are oxidized by O_2 in *oxidative phosphorylation* (Chapter 18).

The citric acid cycle, in conjunction with oxidative phosphorylation, provides the preponderance of energy used by aerobic cells—in human beings, greater than 90%. It is highly efficient because the oxidation of a limited number of citric acid cycle molecules can generate large amounts of NADH and $FADH_2$. Note in Figure 17.2 that the four-carbon molecule, oxaloacetate, that initiates the first step in the citric acid cycle is regenerated at the end of one passage through the cycle. Thus, one molecule of oxaloacetate is capable of participating in the oxidation of many acetyl molecules.

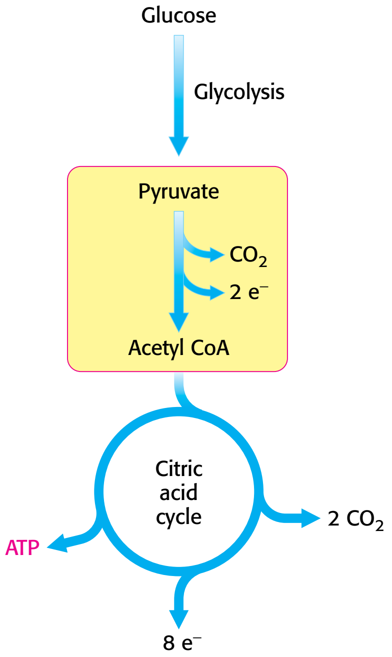


FIGURE 17.4 The link between glycolysis and the citric acid cycle. Pyruvate produced by glycolysis is converted into acetyl CoA, the fuel of the citric acid cycle.

17.1 The Pyruvate Dehydrogenase Complex Links Glycolysis to the Citric Acid Cycle

Carbohydrates, most notably glucose, are processed by glycolysis into pyruvate (Chapter 16). Under anaerobic conditions, the pyruvate is converted into lactate or ethanol, depending on the organism. Under aerobic conditions, the pyruvate is transported into mitochondria by a specific carrier protein embedded in the mitochondrial membrane. In the mitochondrial matrix, pyruvate is oxidatively decarboxylated by the *pyruvate dehydrogenase complex* to form acetyl CoA.

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \longrightarrow \text{acetyl CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+$$

This irreversible reaction is the link between glycolysis and the citric acid cycle (Figure 17.4). Note that the pyruvate dehydrogenase complex produces CO_2 and captures high-transfer-potential electrons in the form of NADH. Thus, the pyruvate dehydrogenase reaction has many of the key features of the reactions of the citric acid cycle itself.

The pyruvate dehydrogenase complex is a large, highly integrated complex of three distinct enzymes (Table 17.1). Pyruvate dehydrogenase complex is a member of a family of homologous complexes that include the citric acid cycle enzyme α-ketoglutarate dehydrogenase complex (p. 505). These complexes are giant, larger than ribosomes, with molecular masses ranging from 4 million to 10 million kDa (Figure 17.5). As we will see, their

TABLE 17.1 Pyruvate dehydrogenase complex of *E. coli*

Enzyme	Abbreviation	Number of chains	Prosthetic group	Reaction catalyzed
Pyruvate dehydrogenase component	E_1	24	TPP	Oxidative decarboxylation of pyruvate
Dihydrolipoyl transacetylase	E_2	24	Lipoamide	Transfer of acetyl group to CoA
Dihydrolipoyl dehydrogenase	E_3	12	FAD	Regeneration of the oxidized form of lipoamide

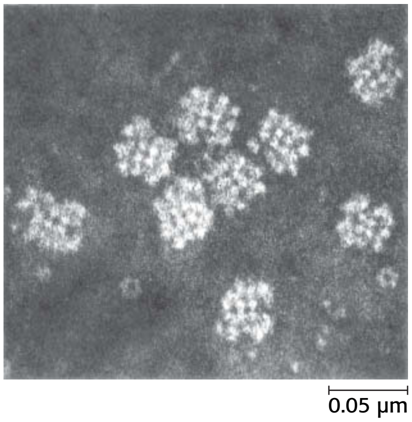
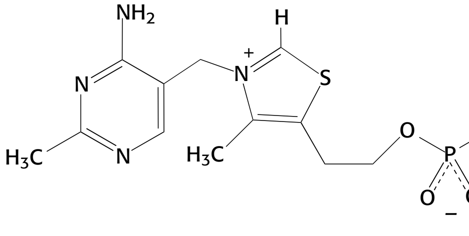


FIGURE 17.5 Electron micrograph of the pyruvate dehydrogenase complex from *E. coli*. [Courtesy of Dr. Lester Reed.]

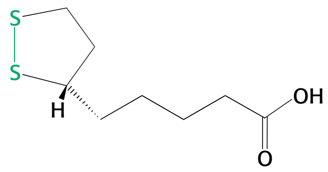
elaborate structures allow groups to travel from one active site to another, connected by tethers to the core of the structure.

Mechanism: The synthesis of acetyl coenzyme A from pyruvate requires three enzymes and five coenzymes

The mechanism of the pyruvate dehydrogenase reaction is wonderfully complex, more so than is suggested by its simple stoichiometry. The reaction requires the participation of the three enzymes of the pyruvate dehydrogenase complex and five coenzymes. The coenzymes *thiamine pyrophosphate* (TPP), *lipoic acid*, and *FAD* serve as catalytic cofactors, and *CoA* and *NAD*$^+$ are stoichiometric cofactors, cofactors that function as substrates.

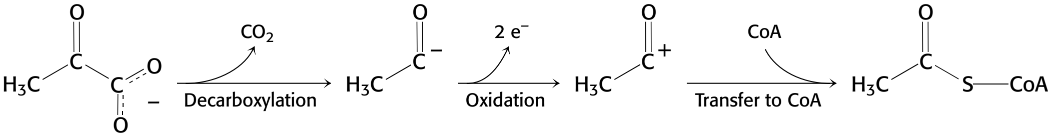


Thiamine pyrophosphate (TPP)



Lipoic acid

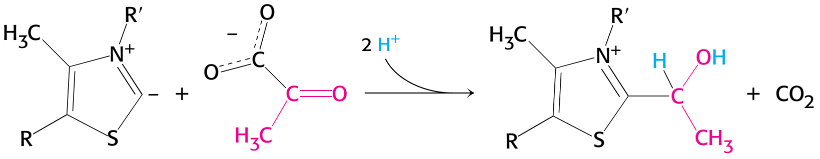
The conversion of pyruvate into acetyl CoA consists of three steps: decarboxylation, oxidation, and transfer of the resultant acetyl group to CoA. A fourth step is required to regenerate the active enzyme.



Pyruvate **Acetyl CoA**

These steps must be coupled to preserve the free energy derived from the decarboxylation step to drive the formation of NADH and acetyl CoA:

1. *Decarboxylation*. Pyruvate combines with TPP and is then decarboxylated to yield hydroxyethyl-TPP (Figure 17.6).



Carbanion of TPP **Pyruvate** **Hydroxyethyl-TPP**

This reaction, the rate limiting step in acetyl CoA synthesis, is catalyzed by the *pyruvate dehydrogenase component* (E_1) of the multienzyme complex. TPP is the prosthetic group of the pyruvate dehydrogenase component.

2. *Oxidation*. The hydroxyethyl group attached to TPP is *oxidized* to form an acetyl group while being simultaneously transferred to lipoamide, a derivative of lipoic acid that is linked to the side chain of a lysine residue by

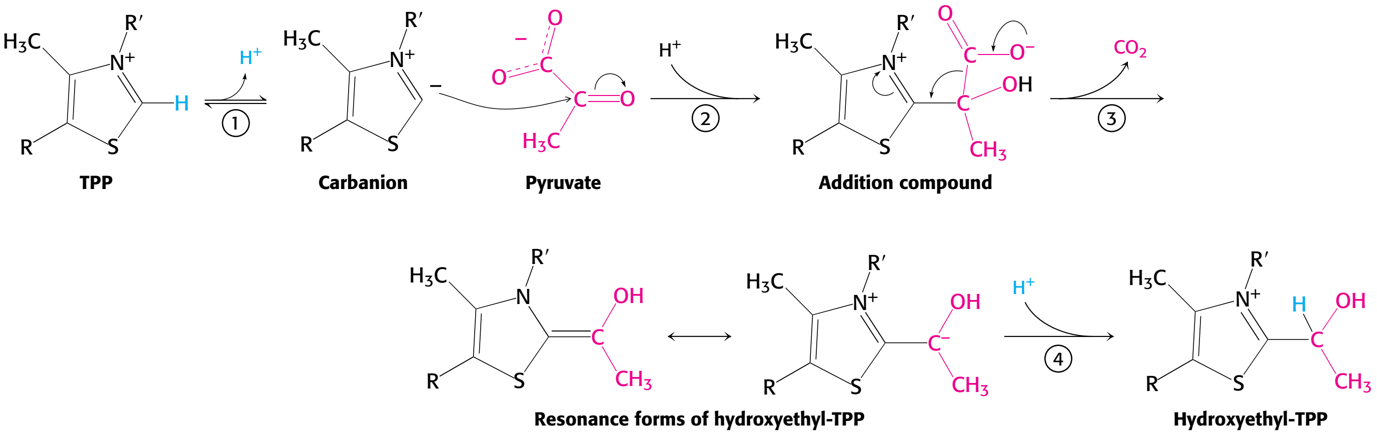
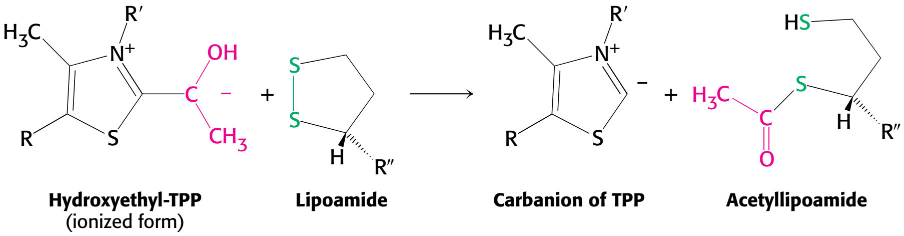

TPP
Carbanion
Pyruvate
Addition compound
Resonance forms of hydroxyethyl-TPP
Hydroxyethyl-TPP
①
②
③
④
H+
CO2

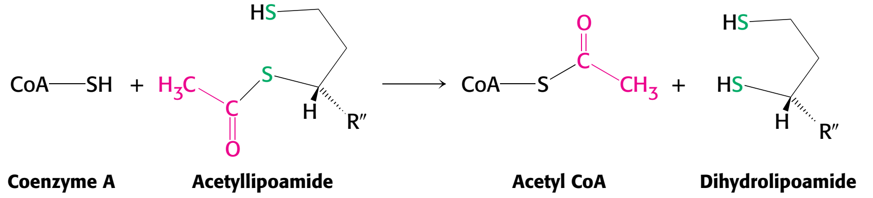
FIGURE 17.6 Mechanism of the E_1 decarboxylation reaction. E_1 is the pyruvate dehydrogenase component of the pyruvate dehydrogenase complex. A key feature of the prosthetic group, TPP, is that the carbon atom between the nitrogen and sulfur atoms in the thiazole ring is much more acidic than most $=\mathrm{C}-$ groups, with a pK_a value near 10. (1) This carbon center ionizes to form a *carbanion.* (2) The carbanion readily adds to the carbonyl group of pyruvate. (3) This addition is followed by the decarboxylation of pyruvate. The positively charged ring of TPP acts as an electron sink that stabilizes the negative charge that is transferred to the ring as part of the decarboxylation. (4) Protonation yields hydroxyethyl-TPP.

an amide linkage. Note that this transfer results in the formation of an energy-rich thioester bond.


Hydroxyethyl-TPP
(ionized form)
Lipoamide
Carbanion of TPP
Acetyllipoamide

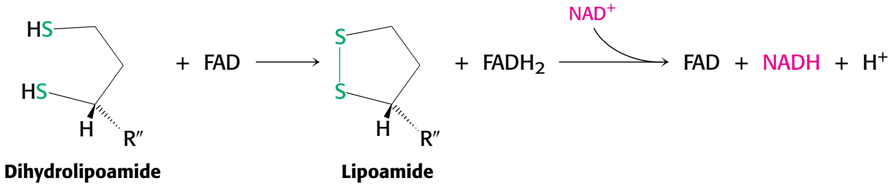
The oxidant in this reaction is the disulfide group of lipoamide, which is reduced to its disulfhydryl form. This reaction, also catalyzed by the pyruvate dehydrogenase component E_1, yields *acetyllipoamide.*

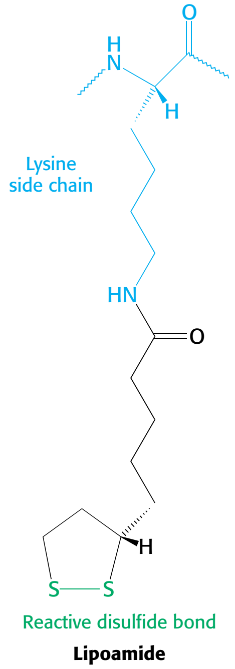
3. *Formation of Acetyl CoA. The acetyl group is transferred from acetyllipoamide to CoA to form acetyl CoA.*


Coenzyme A
Acetyllipoamide
Acetyl CoA
Dihydrolipoamide

Dihydrolipoyl transacetylase (E_2) catalyzes this reaction. The energy-rich thioester bond is preserved as the acetyl group is transferred to CoA. Recall that CoA serves as a carrier of many activated acyl groups, of which acetyl is the simplest (Section 15.3). Acetyl CoA, the fuel for the citric acid cycle, has now been generated from pyruvate.

4. *Regeneration of oxidized lipoamide*. The pyruvate dehydrogenase complex cannot complete another catalytic cycle until the dihydrolipoamide is oxidized to lipoamide. In the fourth step, *the oxidized form of lipoamide is regenerated by dihydrolipoyl dehydrogenase* (E_3). Two electrons are transferred to an FAD prosthetic group of the enzyme and then to NAD^+.


Dihydrolipoamide
Lipoamide
NAD+


Lysine
side chain
Reactive disulfide bond
Lipoamide

This electron transfer from FAD to NAD^+ is unusual because the common role for FAD is to receive electrons from NADH. The electron-transfer potential of FAD is increased by its chemical environment within the enzyme, enabling it to transfer electrons to NAD^+. Proteins tightly associated with FAD or flavin mononucleotide (FMN) are called *flavoproteins.*

Flexible linkages allow lipoamide to move between different active sites

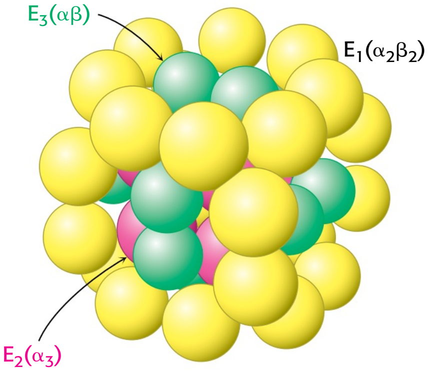


FIGURE 17.7 Schematic representation of the pyruvate dehydrogenase complex. The transacetylase core (E_2) is shown in red, the pyruvate dehydrogenase component (E_1) in yellow, and the dihydrolipoyl dehydrogenase (E_3) in green.

The structures of all of the component enzymes of the bacterial pyruvate dehydrogenase complex are known, albeit from different complexes and species. Thus, it is now possible to construct an atomic model of the complex to understand its activity (Figure 17.7).

The core of the complex is formed by the transacetylase component E_2. Transacetylase consists of eight catalytic trimers assembled to form a hollow cube. Each of the three subunits forming a trimer has three major domains (Figure 17.8). At the amino terminus is a small domain that contains a bound flexible lipoamide cofactor attached to a lysine residue. This domain is homologous to biotin-binding domains such as that of pyruvate carboxylase (Figure 16.26). The lipoamide domain is followed by a small domain that interacts with E_3 within the complex. A larger transacetylase domain completes an E_2 subunit. E_1 is an $\alpha_2\beta_2$ tetramer, and E_3 is an $\alpha\beta$ dimer. Multiple copies of E_1 and E_3 surround the E_2 core. In mammals, this core contains another protein, E_3-binding protein (E_3-BP), which facilitates the interaction between E_2 and E_3. If E_3-BP is missing, the complex has greatly reduced activity.

How do the three distinct active sites work in concert (Figure 17.9)? The key is the long, flexible lipoamide arm of the E_2 subunit, which carries substrate from active site to active site:

1. Pyruvate is decarboxylated at the active site of E_1, forming the hydroxyethyl-TPP intermediate, and CO_2 leaves as the first product. This active site lies deep within the E_1 complex, connected to the enzyme surface by a 20-Å-long hydrophobic channel.

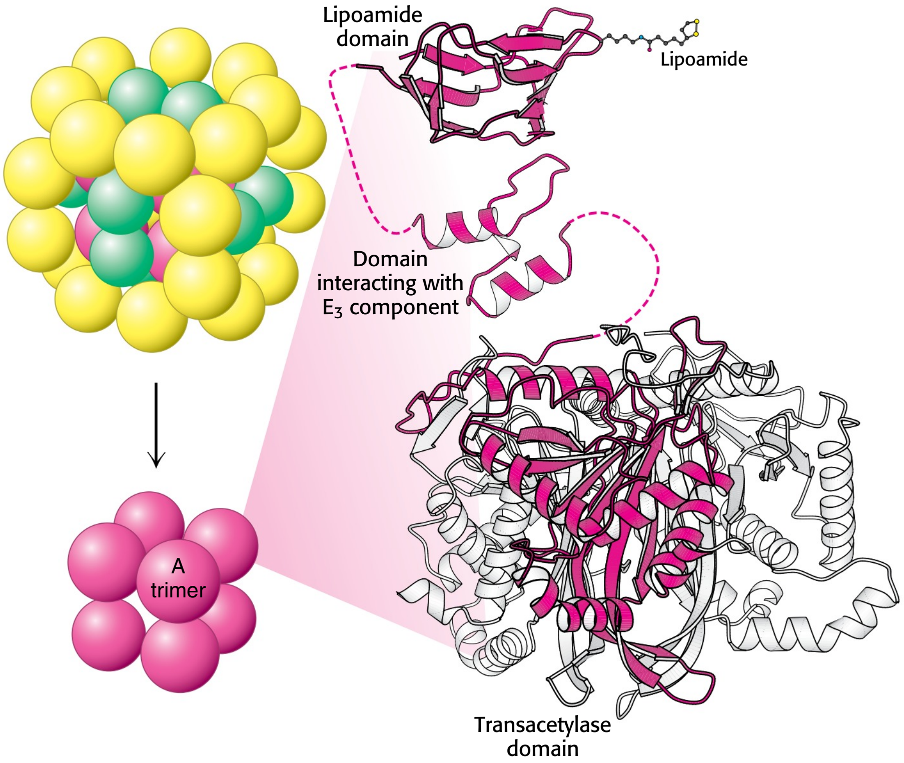


FIGURE 17.8 Structure of the transacetylase (E_2) core. Each red ball represents a trimer of three E_2 subunits. *Notice* that each subunit consists of three domains: a lipoamide-binding domain, a small domain for interaction with E_3, and a large transacetylase catalytic domain. The transacetylase domains of three identical subunits are shown, with one depicted in red and the others in white in the ribbon representation.

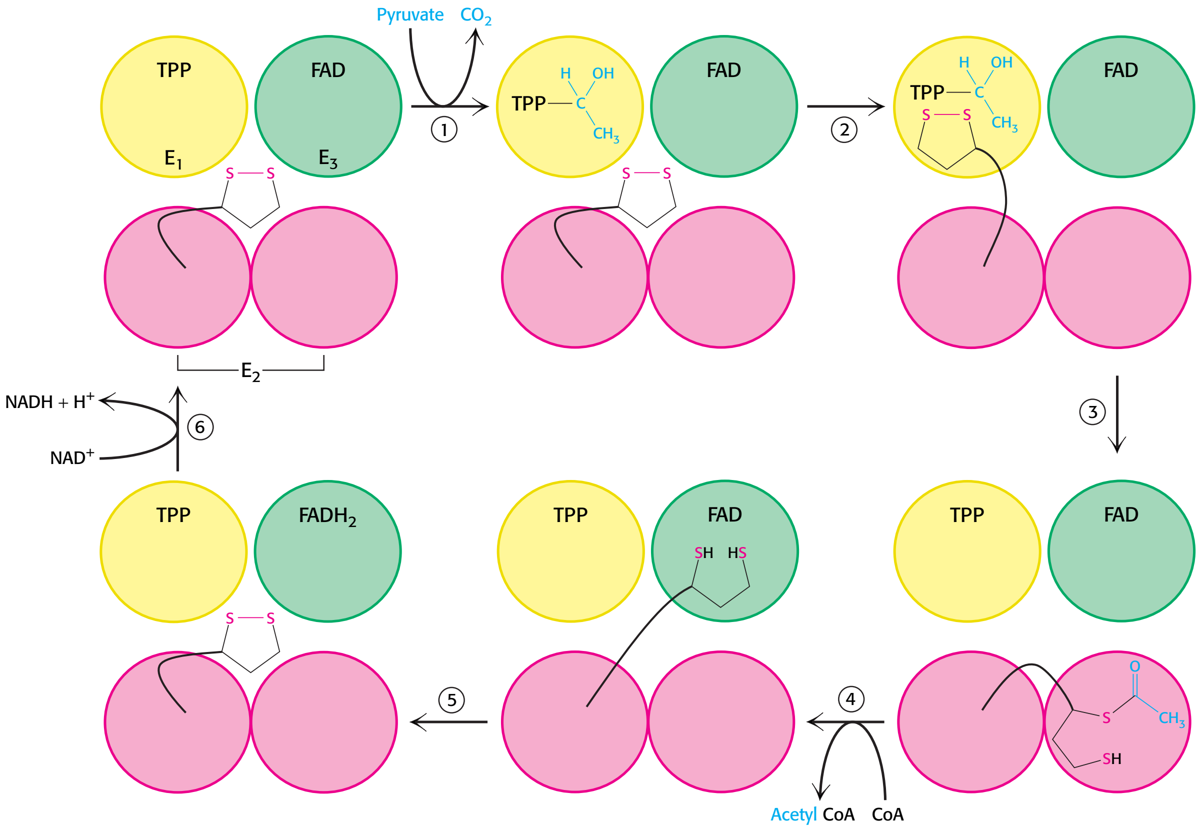


FIGURE 17.9 Reactions of the pyruvate dehydrogenase complex. At the top (left), the enzyme (represented by a yellow, a green, and two red spheres) is unmodified and ready for a catalytic cycle. (1) Pyruvate is decarboxylated to form hydroxyethyl-TPP. (2) The lipoamide arm of E_2 moves into the active site of E_1. (3) E_1 catalyzes the transfer of the two-carbon group to the lipoamide group to form the acetyl–lipoamide complex. (4) E_2 catalyzes the transfer of the acetyl moiety to CoA to form the product acetyl CoA. The dihydrolipoamide arm then swings to the active site of E_3. E_3 catalyzes (5) the oxidation of the dihydrolipoamide and (6) the transfer of the protons and electrons to NAD^+ to complete the reaction cycle.

2. E_2 inserts the lipoamide arm of the lipoamide domain into the deep channel in E_1 leading to the active site.

3. E_1 catalyzes the transfer of the acetyl group to the lipoamide. The acetylated arm then leaves E_1 and enters the E_2 cube to visit the active site of E_2, located deep in the cube at the subunit interface.

4. The acetyl moiety is then transferred to CoA, and the second product, acetyl CoA, leaves the cube. The reduced lipoamide arm then swings to the active site of the E_3 flavoprotein.

5. At the E_3 active site, the lipoamide is oxidized by coenzyme FAD. The reactivated lipoamide is ready to begin another reaction cycle.

6. The final product, NADH, is produced with the reoxidation of $FADH_2$ to FAD.

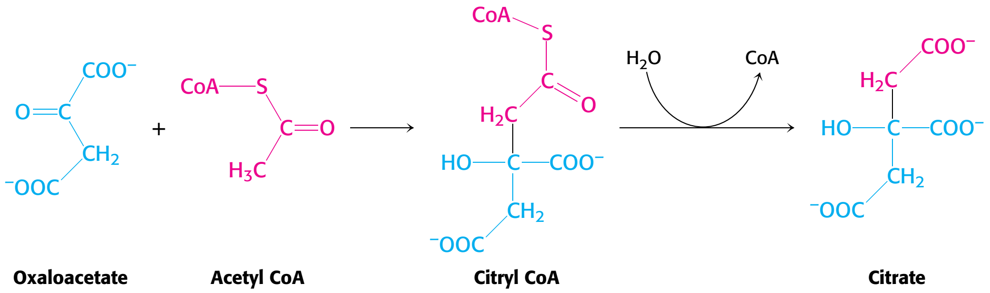
The structural integration of three kinds of enzymes and the long, flexible lipoamide arm make the coordinated catalysis of a complex reaction possible. The proximity of one enzyme to another *increases the overall reaction rate and minimizes side reactions.* All the intermediates in the oxidative decarboxylation of pyruvate remain bound to the complex throughout the reaction sequence and are readily transferred as the flexible arm of E_2 calls on each active site in turn.

17.2 The Citric Acid Cycle Oxidizes Two-Carbon Units

The conversion of pyruvate into acetyl CoA by the pyruvate dehydrogenase complex is the link between glycolysis and cellular respiration because *acetyl CoA is the fuel for the citric acid cycle.* Indeed, all fuels are ultimately metabolized to acetyl CoA or components of the citric acid cycle.

Citrate synthase forms citrate from oxaloacetate and acetyl coenzyme A

The citric acid cycle begins with the condensation of a four-carbon unit, oxaloacetate, and a two-carbon unit, the acetyl group of acetyl CoA. Oxaloacetate reacts with acetyl CoA and H_2O to yield citrate and CoA.



This reaction, which is an aldol condensation followed by a hydrolysis, is catalyzed by *citrate synthase*. Oxaloacetate first condenses with acetyl CoA to form *citryl CoA*, a molecule that is energy rich because it contains the thioester bond that originated in acetyl CoA. The hydrolysis of citryl CoA thioester to citrate and CoA drives the overall reaction far in the direction of the synthesis of citrate. In essence, the hydrolysis of the thioester powers the synthesis of a new molecule from two precursors.

Synthase
An enzyme catalyzing a synthetic reaction in which two units are joined usually without the direct participation of ATP (or another nucleoside triphosphate).

Mechanism: The mechanism of citrate synthase prevents undesirable reactions

Because the condensation of acetyl CoA and oxaloacetate initiates the citric acid cycle, it is very important that side reactions, notably the hydrolysis of acetyl CoA to acetate and CoA, be minimized. Let us briefly consider how the citrate synthase prevents the wasteful hydrolysis of acetyl CoA.

Mammalian citrate synthase is a dimer of identical 49-kDa subunits. Each active site is located in a cleft between the large and the small domains of a subunit, adjacent to the subunit interface. X-ray crystallographic studies of citrate synthase and its complexes with several substrates and inhibitors have revealed that the enzyme undergoes large conformational changes in the course of catalysis. Citrate synthase exhibits sequential, ordered kinetics: oxaloacetate binds first, followed by acetyl CoA. The reason for the ordered binding is that *oxaloacetate induces a major structural rearrangement leading to the creation of a binding site for acetyl CoA*. The binding of oxaloacetate converts the open form of the enzyme into a closed form (Figure 17.10). In each subunit, the small domain rotates 19 degrees relative to the large domain. *Movements as large as 15 Å are produced by the rotation of α helices elicited by quite small shifts of side chains around bound oxaloacetate.* These structural changes create a binding site for acetyl CoA.

Citrate synthase catalyzes the condensation reaction by bringing the substrates into close proximity, orienting them, and polarizing certain bonds (Figure 17.11). The donation and removal of protons transforms acetyl CoA into an *enol intermediate*. The enol attacks oxaloacetate to form a carbon−carbon double bond linking acetyl CoA and oxaloacetate. The newly formed citryl CoA induces additional structural changes in the enzyme, causing the active site to become completely enclosed. The enzyme cleaves the citryl CoA thioester by hydrolysis. CoA leaves the enzyme, followed by citrate, and the enzyme returns to the initial open conformation.

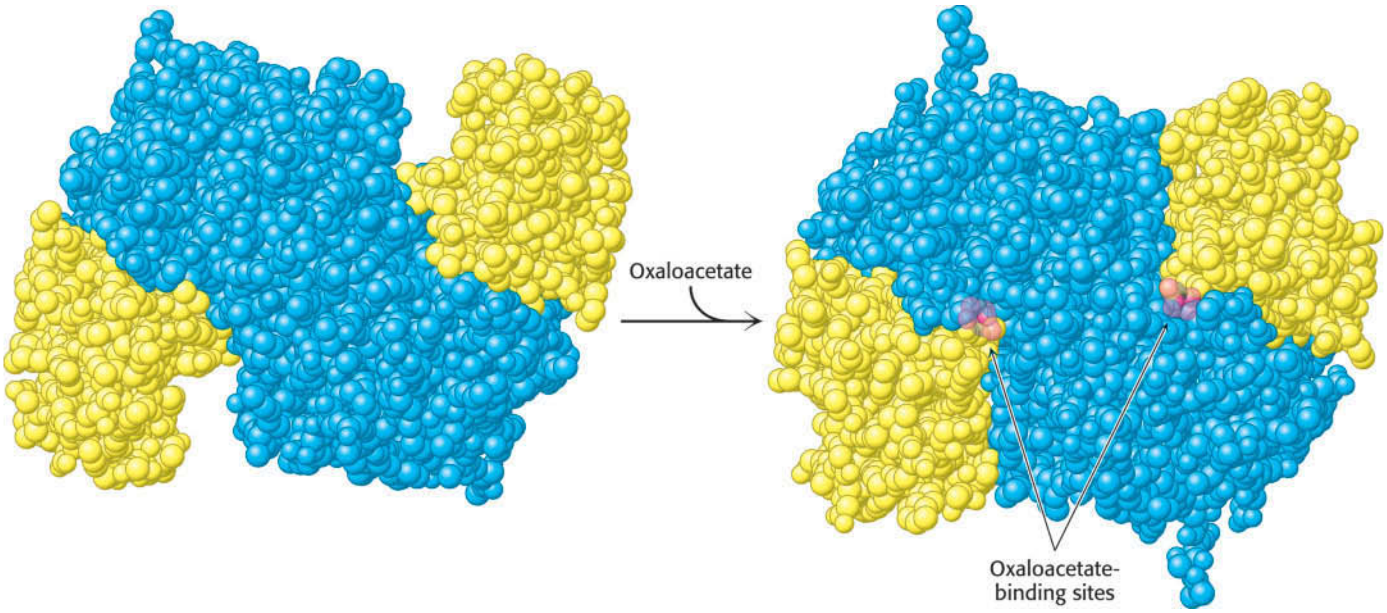
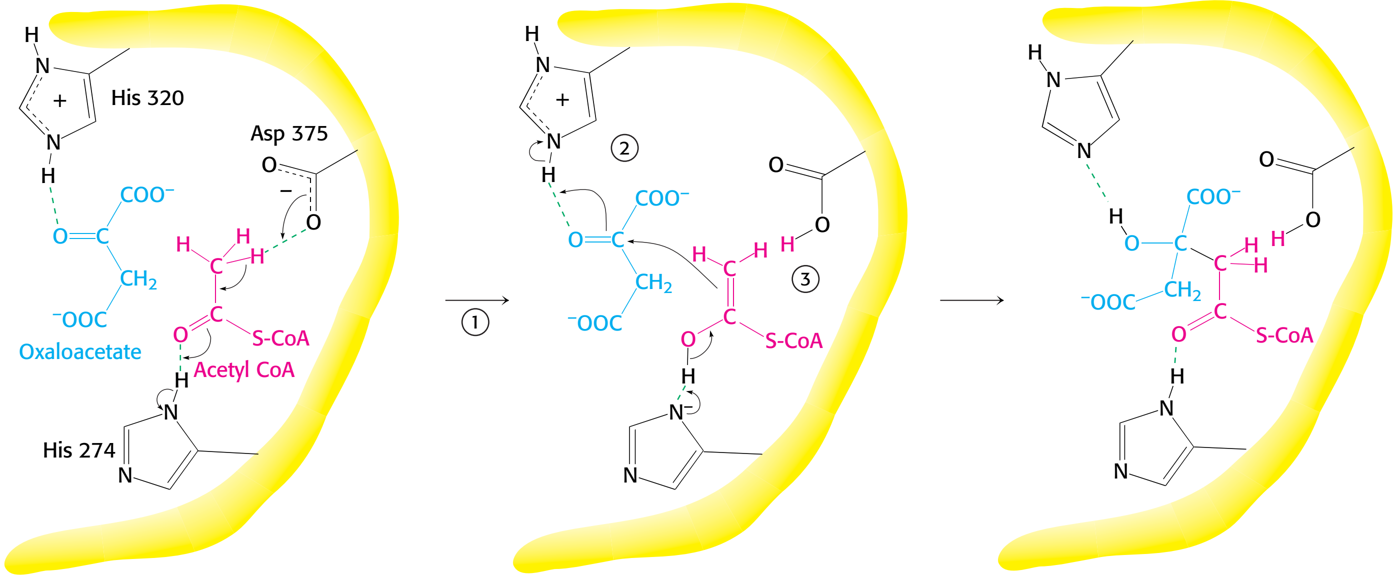


FIGURE 17.10 Conformational changes in citrate synthase on binding oxaloacetate. The small domain of each subunit of the homodimer is shown in yellow; the large domains are shown in blue. (Left) Open form of enzyme alone. (Right) Closed form of the oxaloacetate-enzyme complex. [Drawn from 5CSC.pdb and 4CTS.pdb.]

We can now understand how the wasteful hydrolysis of acetyl CoA is prevented. Citrate synthase is well suited to hydrolyze *citryl* CoA but not *acetyl* CoA. How is this discrimination accomplished? First, acetyl CoA does not bind to the enzyme until oxaloacetate is bound and ready for condensation. Second, the catalytic residues crucial for the hydrolysis of the thioester linkage are not appropriately positioned *until citryl CoA is formed.* As with hexokinase and triose phosphate isomerase (Section 16.1), *induced fit prevents an undesirable side reaction.*

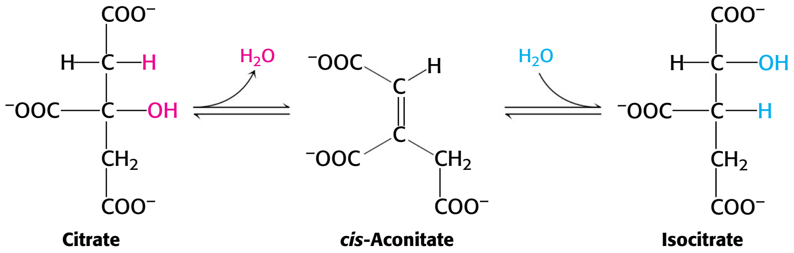


Substrate complex | Enol intermediate | Citryl CoA complex

FIGURE 17.11 Mechanism of synthesis of citryl CoA by citrate synthase. (1) In the substrate complex (left), His 274 donates a proton to the carbonyl oxygen of acetyl CoA to promote the removal of a methyl proton by Asp 375 to form the enol intermediate (center). (2) Oxaloacetate is activated by the transfer of a proton from His 320 to its carbonyl carbon atom. (3) Simultaneously, the enol of acetyl CoA attacks the carbonyl carbon of oxaloacetate to form a carbon–carbon bond linking acetyl CoA and oxaloacetate. His 274 is reprotonated. Citryl CoA is formed. His 274 participates again as a proton donor to hydrolyze the thioester (not shown), yielding citrate and CoA.

Citrate is isomerized into isocitrate

The hydroxyl group is not properly located in the citrate molecule for the oxidative decarboxylations that follow (Problem 27). Thus, citrate is isomerized into isocitrate to enable the six-carbon unit to undergo oxidative decarboxylation. The isomerization of citrate is accomplished by a *dehydration* step followed by a *hydration* step. The result is an interchange of an H and an OH. The enzyme catalyzing both steps is called *aconitase* because cis-*aconitate* is an intermediate.



Aconitase is an *iron–sulfur protein,* or *nonheme-iron protein,* in that it contains iron that is not bonded to heme. Rather, its four iron atoms are complexed to four inorganic sulfides and three cysteine sulfur atoms, leaving one iron atom available to bind citrate through one of its COO^- groups and an OH group (Figure 17.12). This Fe-S cluster participates in dehydrating and rehydrating the bound substrate.

Isocitrate is oxidized and decarboxylated to alpha-ketoglutarate

We come now to the first of four oxidation–reduction reactions in the citric acid cycle. The oxidative decarboxylation of isocitrate is catalyzed by *isocitrate dehydrogenase.*

$$\text{Isocitrate} + \text{NAD}^+ \longrightarrow \alpha\text{-ketoglutarate} + \text{CO}_2 + \text{NADH}$$

The intermediate in this reaction is oxalosuccinate, an unstable β-ketoacid. While bound to the enzyme, it loses CO_2 to form α-ketoglutarate.

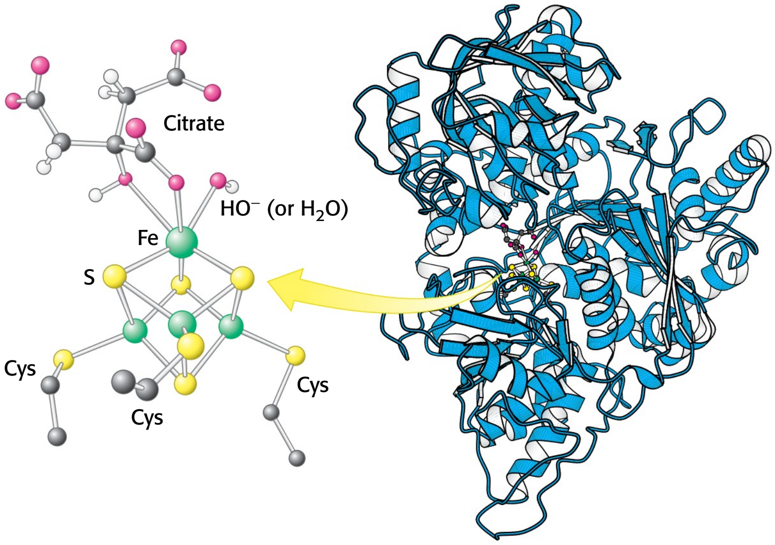
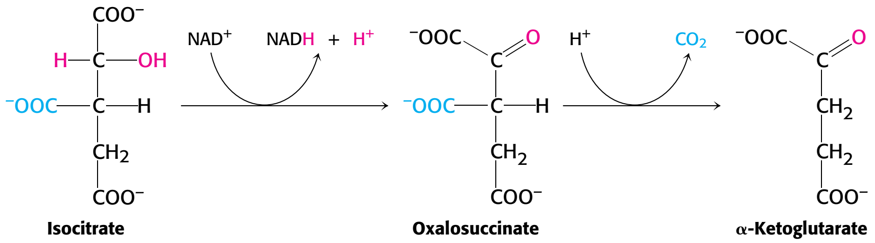


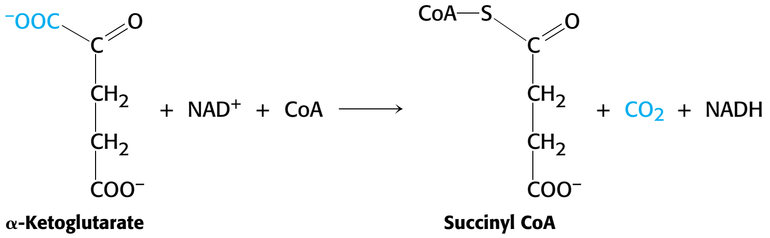
FIGURE 17.12 Binding of citrate to the iron–sulfur complex of aconitase. A 4Fe-4S iron–sulfur cluster is a component of the active site of aconitase. *Notice* that one of the iron atoms of the cluster binds to a COO^- group and an OH group of citrate. [Drawn from 1C96.pdb.]



This oxidation generates the first high-transfer-potential electron carrier, NADH, in the cycle.

Succinyl coenzyme A is formed by the oxidative decarboxylation of alpha-ketoglutarate

The conversion of isocitrate into α-ketoglutarate is followed by a second oxidative decarboxylation reaction, the formation of succinyl CoA from α-ketoglutarate.



This reaction is catalyzed by the α-*ketoglutarate dehydrogenase complex,* an organized assembly of three kinds of enzymes that is homologous to the pyruvate dehydrogenase complex. In fact, the E_3 component is identical in both enzymes. The oxidative decarboxylation of α-ketoglutarate closely resembles that of pyruvate, also an α-ketoacid.

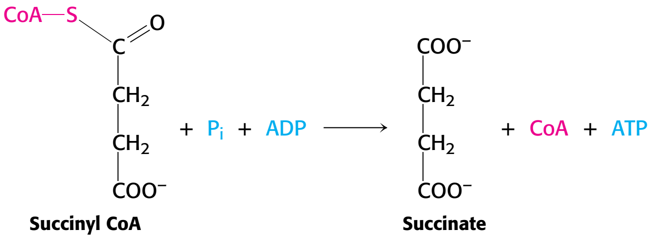
$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \xrightarrow{\text{Pyruvate dehydrogenase complex}} \text{acetyl CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+$$

$$\alpha\text{-Ketoglutarate} + \text{CoA} + \text{NAD}^+ \xrightarrow{\alpha\text{-Ketoglutarate dehydrogenase complex}} \text{succinyl CoA} + \text{CO}_2 + \text{NADH}$$

Both reactions include the decarboxylation of an α-ketoacid and the subsequent formation of a thioester linkage with CoA that has a high transfer potential. The reaction mechanisms are entirely analogous (p. 497).

A compound with high phosphoryl-transfer potential is generated from succinyl coenzyme A

Succinyl CoA is an energy-rich thioester compound. The $\Delta G^{\circ\prime}$ for the hydrolysis of succinyl CoA is about -33.5 kJ mol^{-1} (-8.0 kcal mol^{-1}), which is comparable to that of ATP (-30.5 kJ mol^{-1}, or -7.3 kcal mol^{-1}). In the citrate synthase reaction, the cleavage of the thioester bond powers the synthesis of the six-carbon citrate from the four-carbon oxaloacetate and the two-carbon fragment. *The cleavage of the thioester bond of succinyl CoA is coupled to the phosphorylation of a purine nucleoside diphosphate, usually ADP.* This reaction, which is readily reversible, is catalyzed by *succinyl CoA synthetase* (succinate thiokinase).


CoA—S
O
C
CH₂
CH₂
COO⁻
Succinyl CoA
+ Pᵢ + ADP
COO⁻
CH₂
CH₂
COO⁻
Succinate
+ CoA + ATP

This reaction is the only step in the citric acid cycle that directly yields a compound with high phosphoryl-transfer potential. In mammals, there are two isozymic forms of the enzyme, one specific for ADP and one for GDP. In tissues that perform large amounts of cellular respiration, such as skeletal and heart muscle, the ADP-requiring isozyme predominates. In tissues that perform many anabolic reactions, such as the liver, the GDP-requiring enzyme is common. The GDP-requiring enzyme is believed to work in reverse of the direction observed in the TCA cycle; that is, GTP is used to power the synthesis of succinyl CoA, which is a precursor for heme synthesis. The *E. coli* enzyme uses either GDP or ADP as the phosphoryl-group acceptor.

Note that the enzyme *nucleoside diphosphokinase,* which catalyzes the following reaction,

$$\text{GTP} + \text{ADP} \rightleftharpoons \text{GDP} + \text{ATP}$$

allows the γ phosphoryl group to be readily transferred from GTP to form ATP, thereby allowing the adjustment of the concentration of GTP or ATP to meet the cell's need.

Mechanism: Succinyl coenzyme A synthetase transforms types of biochemical energy

The mechanism of this reaction is a clear example of an energy transformation: energy inherent in the thioester molecule is transformed into phosphoryl-group-transfer potential (Figure 17.13). The first step is the displacement of coenzyme A by orthophosphate, which generates

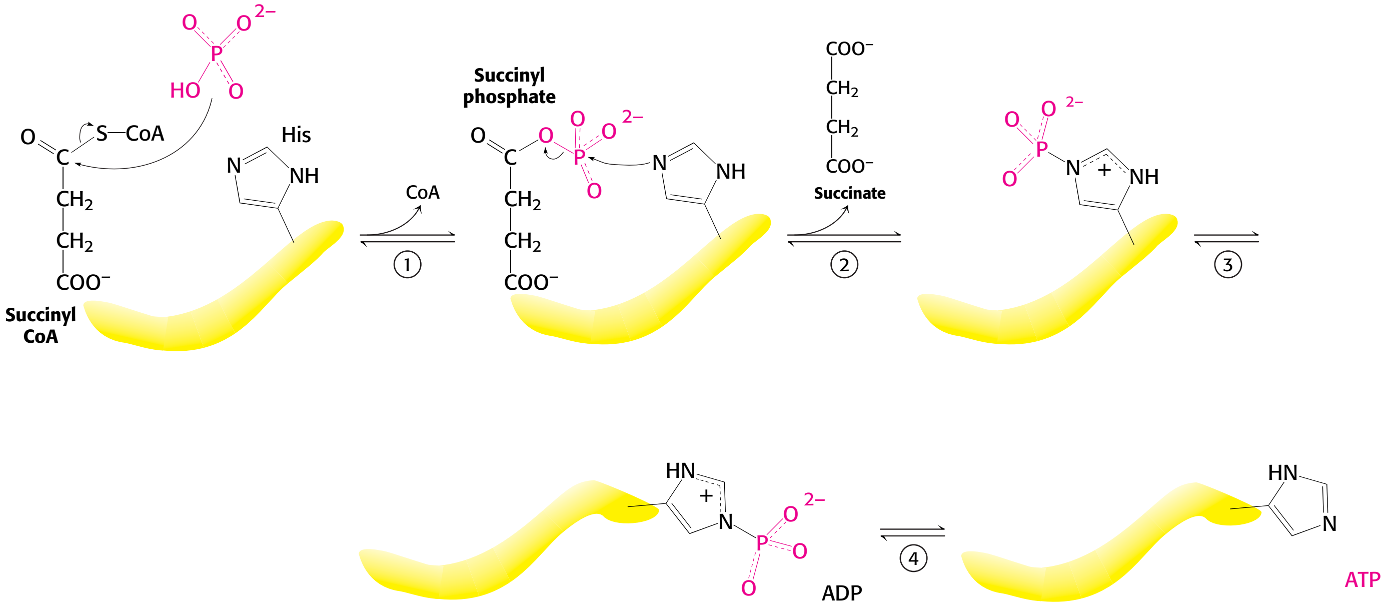

O
O
2−
P
HO
O
O
C
S—CoA
CH₂
CH₂
COO⁻
Succinyl CoA
His
N
NH
CoA
①
Succinyl phosphate
O
C
O
O
P
2−
O
CH₂
CH₂
COO⁻
N
NH
COO⁻
CH₂
CH₂
COO⁻
Succinate
②
O
O
2−
P
O
N
+
NH
③
HN
+
N
O
2−
P
O
O
ADP
④
HN
N
ATP

FIGURE 17.13 Reaction mechanism of succinyl CoA synthetase. The reaction proceeds through a phosphorylated enzyme intermediate. (1) Orthophosphate displaces coenzyme A, which generates another energy-rich compound, succinyl phosphate. (2) A histidine residue removes the phosphoryl group with the concomitant generation of succinate and phosphohistidine. (3) The phosphohistidine residue then swings over to a bound ADP, and (4) the phosphoryl group is transferred to form ATP.

another energy-rich compound, succinyl phosphate. A histidine residue plays a key role as a moving arm that detaches the phosphoryl group, then swings over to a bound ADP and transfers the group to form ATP. The participation of high-energy compounds in all the steps is attested to by the fact that the reaction is readily reversible: $\Delta G^{\circ\prime} = -3.4 \text{ kJ mol}^{-1}$ $(-0.8 \text{ kcal mol}^{-1})$. The formation of ATP at the expense of succinyl CoA is an example of substrate-level phosphorylation (Section 16.1).

Succinyl CoA synthetase is an $\alpha_2\beta_2$ heterodimer; the functional unit is one $\alpha\beta$ pair. The enzyme mechanism shows that a phosphoryl group is transferred first to succinyl CoA bound in the α subunit and then to a nucleoside diphosphate bound in the β subunit. Examination of the three-dimensional structure of succinyl CoA synthetase reveals that each subunit comprises two domains (Figure 17.14). The amino-terminal domains of the two subunits have different structures, each characteristic of its role in the mechanism. The amino-terminal domain of the α subunit forms a Rossmann fold (Section 16.1), which binds the ADP substrate of succinyl CoA synthetase. The amino-terminal domain of the β subunit is an ATP-grasp domain, found in many enzymes, which here binds and activates ADP. Succinyl CoA synthetase has evolved by adopting these domains and harnessing them to capture the energy associated with succinyl CoA cleavage, which is used to drive the generation of a nucleoside triphosphate.

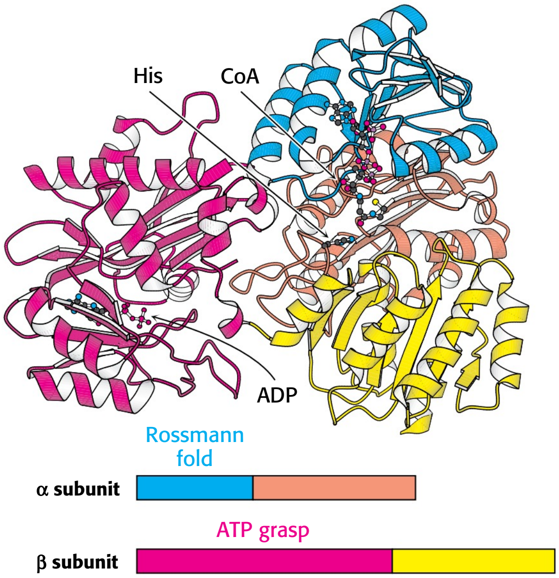
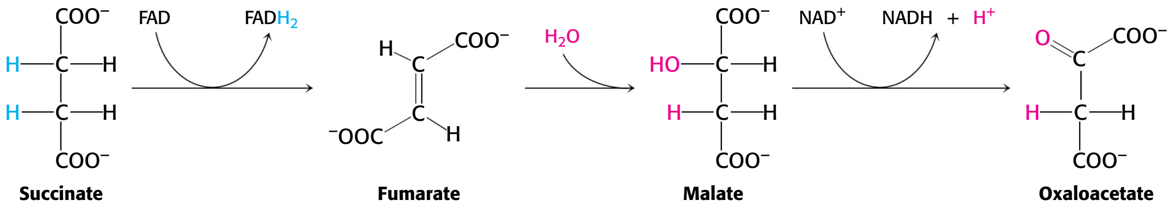


FIGURE 17.14 Structure of succinyl CoA synthetase. The enzyme is composed of two subunits. The α subunit contains a Rossmann fold that binds the ADP component of CoA, and the β subunit contains a nucleotide-activating region called the ATP-grasp domain. The ATP-grasp domain is shown here binding a molecule of ADP. *Notice* that the histidine residue is between the CoA and the ADP. This histidine residue picks up the phosphoryl group from near the CoA and swings over to transfer it to the nucleotide bound in the ATP-grasp domain. [Drawn from 1CQI.pdb.]

Oxaloacetate is regenerated by the oxidation of succinate

Reactions of four-carbon compounds constitute the final stage of the citric acid cycle: the regeneration of oxaloacetate.



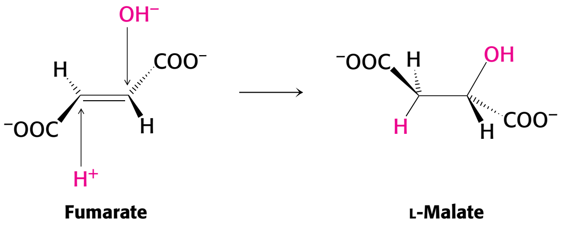
The reactions constitute a metabolic motif that we will see again in fatty acid synthesis and degradation as well as in the degradation of some amino acids. A methylene group (CH_2) is converted into a carbonyl group C=O in three steps: an oxidation, a hydration, and a second oxidation reaction. Oxaloacetate is thereby regenerated for another round of the cycle, and more energy is extracted in the form of $FADH_2$ and NADH.

Succinate is oxidized to fumarate by *succinate dehydrogenase.* The hydrogen acceptor is FAD rather than NAD^+, which is used in the other three oxidation reactions in the cycle. FAD is the hydrogen acceptor in this reaction because the free-energy change is insufficient to reduce NAD^+. FAD is nearly always the electron acceptor in oxidations that remove two hydrogen *atoms* from a substrate. In succinate dehydrogenase, the isoalloxazine ring of FAD is covalently attached to a histidine side chain of the enzyme (denoted E-FAD).

$$\text{E-FAD} + \text{succinate} \rightleftharpoons \text{E-FADH}_2 + \text{fumarate}$$

Succinate dehydrogenase, like aconitase, is an iron–sulfur protein. Indeed, succinate dehydrogenase contains three different kinds of iron–sulfur clusters: 2Fe-2S (two iron atoms bonded to two inorganic sulfides), 3Fe-4S, and 4Fe-4S. Succinate dehydrogenase—which consists of a 70-kDa and a 27-kDa subunit—differs from other enzymes in the citric acid cycle in being embedded in the inner mitochondrial membrane. In fact, *succinate dehydrogenase is directly associated with the electron-transport chain, the link between the citric acid cycle and ATP formation.* $FADH_2$ produced by the oxidation of succinate does not dissociate from the enzyme, in contrast with NADH produced in other oxidation–reduction reactions. Rather, two electrons are transferred from $FADH_2$ directly to iron–sulfur clusters of the enzyme, which in turn passes the electrons to coenzyme Q (CoQ). Coenzyme Q, an important member of the electron-transport chain, passes electrons to the ultimate acceptor, molecular oxygen, as we shall see in Chapter 18.

The next step is the hydration of fumarate to form L-malate. *Fumarase* catalyzes a stereospecific trans addition of H^+ and OH^-. The OH^- group adds to only one side of the double bond of fumarate; hence, only the L isomer of malate is formed.



Finally, malate is oxidized to form oxaloacetate. This reaction is catalyzed by *malate dehydrogenase,* and NAD^+ is again the hydrogen acceptor.

$$\text{Malate} + NAD^+ \rightleftharpoons \text{oxaloacetate} + NADH + H^+$$

The standard free energy for this reaction, unlike that for the other steps in the citric acid cycle, is significantly positive ($\Delta G^{\circ\prime} = +29.7 \text{ kJ mol}^{-1}$, or $+7.1 \text{ kcal mol}^{-1}$). The oxidation of malate is driven by the use of the products—oxaloacetate by citrate synthase and NADH by the electron-transport chain.

The citric acid cycle produces high-transfer-potential electrons, ATP, and CO_2

The net reaction of the citric acid cycle is

$$\text{Acetyl CoA} + 3\,NAD^+ + FAD + ADP + P_i + 2\,H_2O \longrightarrow$$
$$2\,CO_2 + 3\,NADH + FADH_2 + ATP + 2\,H^+ + CoA$$

Let us recapitulate the reactions that give this stoichiometry (Figure 17.15 and Table 17.2):

1. Two carbon atoms enter the cycle in the condensation of an acetyl unit (from acetyl CoA) with oxaloacetate. Two carbon atoms leave the cycle in the form of CO_2 in the successive decarboxylations catalyzed by isocitrate dehydrogenase and α-ketoglutarate dehydrogenase.

2. Four pairs of hydrogen atoms leave the cycle in four oxidation reactions. Two NAD^+ molecules are reduced in the oxidative decarboxylations of isocitrate and α-ketoglutarate, one FAD molecule is reduced in the oxidation

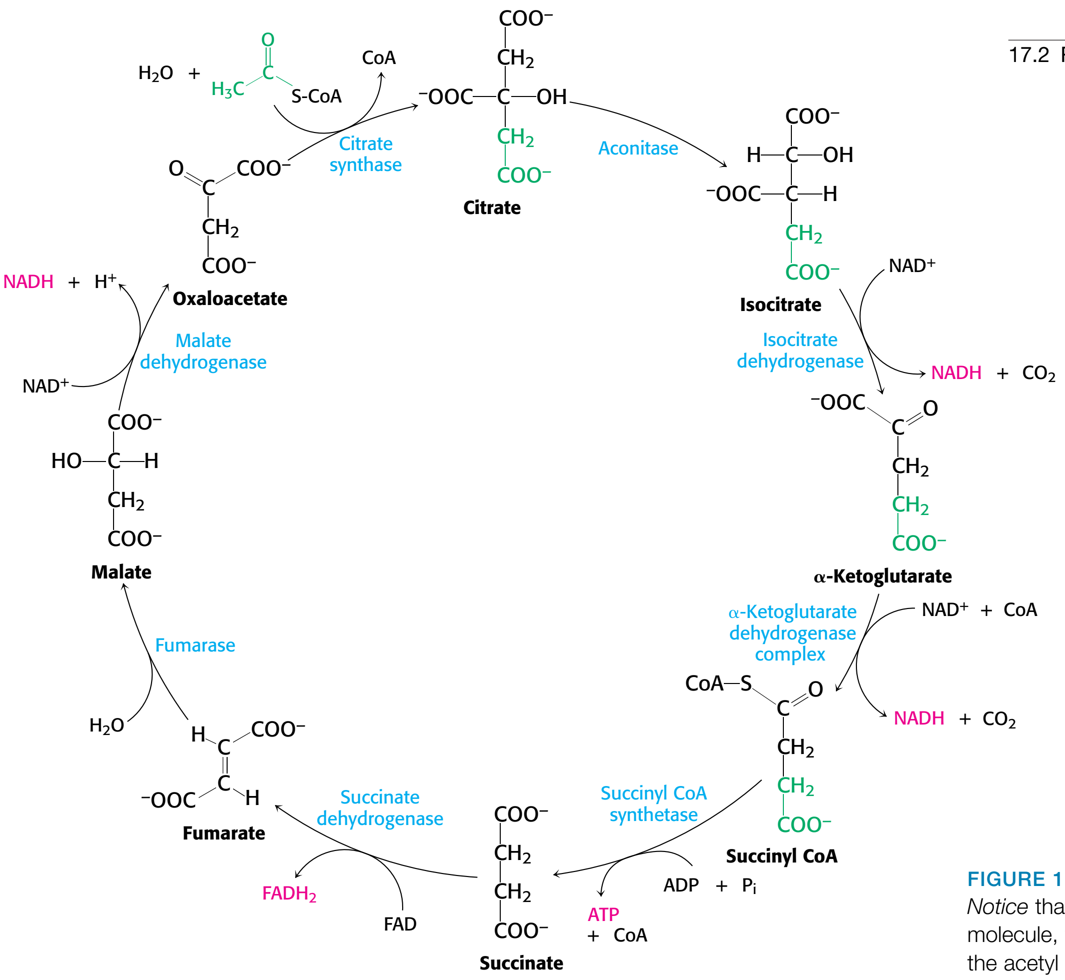

H2O +
H3C–C(=O)–S-CoA
CoA
Citrate synthase
Citrate
Aconitase
Isocitrate
NAD+
Isocitrate dehydrogenase
NADH + CO2
α-Ketoglutarate
NAD+ + CoA
α-Ketoglutarate dehydrogenase complex
NADH + CO2
Succinyl CoA
Succinyl CoA synthetase
ADP + Pi
ATP + CoA
Succinate
Succinate dehydrogenase
FAD
FADH2
Fumarate
Fumarase
H2O
Malate
Malate dehydrogenase
NAD+
NADH + H+
Oxaloacetate

FIGURE 17.15 The citric acid cycle. *Notice* that since succinate is a symmetric molecule, the identity of the carbons from the acetyl unit is lost.

TABLE 17.2 Citric acid cycle

Step	Reaction	Enzyme	Prosthetic group	Type*	$\Delta G^{\circ\prime}$ kJ mol^{-1}	$\Delta G^{\circ\prime}$ kcal mol^{-1}
1	Acetyl CoA + oxaloacetate + $H_2O \rightarrow$ citrate + CoA + H^+	Citrate synthase		a	–31.4	–7.5
2a	Citrate $\rightleftharpoons$ *cis*-aconitate + H_2O	Aconitase	Fe-S	b	+8.4	+2.0
2b	*cis*-Aconitate + $H_2O \rightleftharpoons$ isocitrate	Aconitase	Fe-S	c	−2.1	−0.5
3	Isocitrate + $NAD^+ \rightleftharpoons$ α-ketoglutarate + CO_2 + NADH	Isocitrate dehydrogenase		d + e	−8.4	−2.0
4	α-Ketoglutarate + NAD^+ + CoA $\rightleftharpoons$ succinyl CoA + CO_2 + NADH	α-Ketoglutarate dehydrogenase complex	Lipoic acid, FAD, TPP	d + e	−30.1	−7.2
5	Succinyl CoA + P_i + ADP $\rightleftharpoons$ succinate + ATP + CoA	Succinyl CoA synthetase		f	−3.3	−0.8
6	Succinate + FAD (enzyme-bound) $\rightleftharpoons$ fumarate + $FADH_2$(enzyme-bound)	Succinate dehydrogenase	FAD, Fe-S	e	0	0
7	Fumarate + $H_2O \rightleftharpoons$ L-malate	Fumarase		c	−3.8	−0.9
8	L-Malate + $NAD^+ \rightleftharpoons$ oxaloacetate + NADH + H^+	Malate dehydrogenase		e	+29.7	+7.1

*Reaction type: (a) condensation; (b) dehydration; (c) hydration; (d) decarboxylation; (e) oxidation; (f) substrate-level phosphorylation.

of succinate, and one NAD^+ molecule is reduced in the oxidation of malate. Recall also that one NAD^+ molecule is reduced in the oxidative decarboxylation of pyruvate to form acetyl CoA.

3. One compound with high phosphoryl-transfer potential, usually ATP, is generated from the cleavage of the thioester linkage in succinyl CoA.

4. Two water molecules are consumed: one in the synthesis of citrate by the hydrolysis of citryl CoA and the other in the hydration of fumarate.

Isotope-labeling studies have revealed that the two carbon atoms that enter each cycle are not the ones that leave. The two carbon atoms that enter the cycle as the acetyl group are retained during the initial two decarboxylation reactions (Figure 17.15) and then remain incorporated in the four-carbon acids of the cycle. Note that succinate is a symmetric molecule. Consequently, the two carbon atoms that enter the cycle can occupy any of the carbon positions in the subsequent metabolism of the four-carbon acids. The two carbons that enter the cycle as the acetyl group will be released as CO_2 in *subsequent* trips through the cycle. To understand why citrate is not processed as a symmetric molecule, see Problems 34 and 35.

Evidence is accumulating that the enzymes of the citric acid cycle are physically associated with one another. The close arrangement of enzymes enhances the efficiency of the citric acid cycle because a reaction product can pass directly from one active site to the next through connecting channels, a process called *substrate channeling*.

As will be considered in Chapter 18, the electron-transport chain oxidizes the NADH and $FADH_2$ formed in the citric acid cycle. The transfer of electrons from these carriers to O_2, the ultimate electron acceptor, leads to the generation of a proton gradient across the inner mitochondrial membrane. This proton-motive force then powers the generation of ATP; the net stoichiometry is about 2.5 ATP per NADH, and 1.5 ATP per $FADH_2$. Consequently, nine high-transfer-potential phosphoryl groups are generated when the electron-transport chain oxidizes 3 NADH molecules and 1 $FADH_2$ molecule, and one ATP is directly formed in one round of the citric acid cycle. Thus, one acetyl unit generates approximately 10 molecules of ATP. In dramatic contrast, the anaerobic glycolysis of an entire glucose molecule generates only 2 molecules of ATP (and 2 molecules of lactate).

Recall that molecular oxygen does not participate directly in the citric acid cycle. However, the cycle operates only under aerobic conditions because NAD^+ and FAD can be regenerated in the mitochondrion only by the transfer of electrons to molecular oxygen. *Glycolysis has both an aerobic and an anaerobic mode, whereas the citric acid cycle is strictly aerobic.* Glycolysis can proceed under anaerobic conditions because NAD^+ is regenerated in the conversion of pyruvate into lactate or ethanol.

17.3 Entry to the Citric Acid Cycle and Metabolism Through It Are Controlled

The citric acid cycle is the final common pathway for the aerobic oxidation of fuel molecules. Moreover, as we will see shortly (Section 17.4) and repeatedly elsewhere in our study of biochemistry, the cycle is an important source of building blocks for a host of important biomolecules. As befits its role as the metabolic hub of the cell, entry into the cycle and the rate of the cycle itself are controlled at several stages.

The pyruvate dehydrogenase complex is regulated allosterically and by reversible phosphorylation

As stated earlier, glucose can be formed from pyruvate (Section 16.3). *However, the formation of acetyl CoA from pyruvate is an irreversible step in animals and thus they are unable to convert acetyl CoA back into glucose.* The oxidative decarboxylation of pyruvate to acetyl CoA commits the carbon atoms of glucose to one of two principal fates: oxidation to CO_2 by the citric acid cycle, with the concomitant generation of energy, or incorporation into lipid (Figure 17.16). As expected of an enzyme at a critical branch point in metabolism, the activity of the pyruvate dehydrogenase complex is stringently controlled. High concentrations of reaction products inhibit the reaction: acetyl CoA inhibits the transacetylase component (E_2) by binding directly, whereas NADH inhibits the dihydrolipoyl dehydrogenase (E_3). High concentrations of NADH and acetyl CoA inform the enzyme that the energy needs of the cell have been met or that fatty acids are being degraded to produce acetyl CoA and NADH. In either case, there is no need to metabolize pyruvate to acetyl CoA. This inhibition has the effect of sparing glucose, because most pyruvate is derived from glucose by glycolysis (Section 16.1).

The key means of regulation of the complex in eukaryotes is covalent modification (Figure 17.17). *Phosphorylation of the pyruvate dehydrogenase component (E_1) by pyruvate dehydrogenase kinase* (PDK) *switches off the activity of the complex.* There are four isozymes of PDK that are expressed in a tissue-specific manner. *Deactivation is reversed by the pyruvate dehydrogenase phosphatase* (PDP), of which there are two isozymic forms. In mammals, the kinase and the phosphatase are associated with the E_2-E_3-BP complex, again highlighting the structural and mechanistic importance of this core. Both the kinase and the phosphatase are regulated. To see how this regulation works in biological conditions, consider muscle that is becoming active after a period of rest (Figure 17.18). At rest, the muscle will not have significant energy demands. Consequently, the $NADH/NAD^+$, acetyl CoA/CoA, and ATP/ADP ratios will be high. These high ratios promote phosphorylation and inactivation of the complex by activating PDK. In other words, high concentrations of immediate (acetyl CoA and NADH) and ultimate (ATP) products inhibit the activity. Thus, *pyruvate dehydrogenase is switched off when the energy charge is high.*

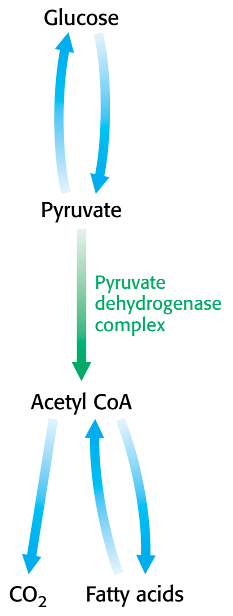


FIGURE 17.16 From glucose to acetyl CoA. The synthesis of acetyl CoA by the pyruvate dehydrogenase complex is a key irreversible step in the metabolism of glucose.

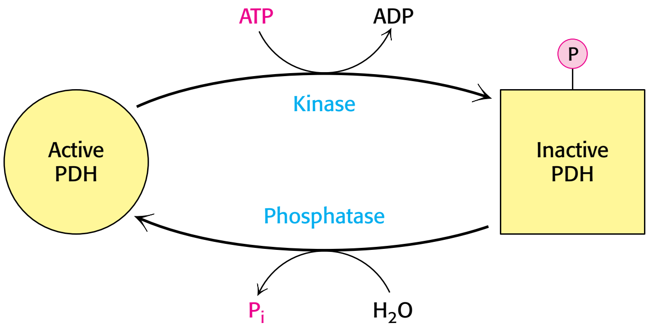


FIGURE 17.17 Regulation of the pyruvate dehydrogenase complex. A specific kinase phosphorylates and inactivates pyruvate dehydrogenase (PDH), and a phosphatase activates the dehydrogenase by removing the phosphoryl group. The kinase and the phosphatase also are highly regulated enzymes.

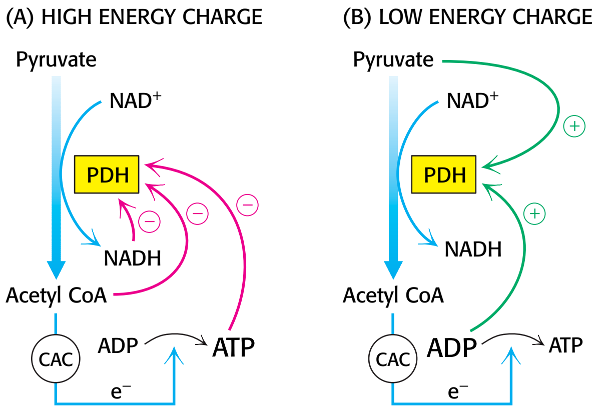


FIGURE 17.18 Response of the pyruvate dehydrogenase complex to the energy charge. The pyruvate dehydrogenase complex is regulated to respond to the energy charge of the cell. (A) The complex is inhibited by its immediate products, NADH and acetyl CoA, as well as by the ultimate product of cellular respiration, ATP. (B) The complex is activated by pyruvate and ADP, which inhibit the kinase that phosphorylates PDH.

As exercise begins, the concentrations of ADP and pyruvate will increase as muscle contraction consumes ATP and glucose is converted into pyruvate to meet the energy demands. Both ADP and pyruvate activate the dehydrogenase by inhibiting the kinase. Moreover, the phosphatase is stimulated by Ca^{2+}, the same signal that initiates muscle contraction. A rise in the cytoplasmic Ca^{2+} level (Section 35.2) elevates the mitochondrial Ca^{2+} level. The rise in mitochondrial Ca^{2+} activates the phosphatase, enhancing pyruvate dehydrogenase activity.

In some tissues, the phosphatase is regulated by hormones. In liver, epinephrine binds to the α-adrenergic receptor to initiate the phosphatidylinositol pathway (Section 14.1), causing an increase in Ca^{2+} concentration that activates the phosphatase. In tissues capable of fatty acid synthesis, such as the liver and adipose tissue, insulin, the hormone that signifies the fed state, stimulates the phosphatase, increasing the conversion of pyruvate into acetyl CoA. Acetyl CoA is the precursor for fatty acid synthesis (Section 22.4). In these tissues, the pyruvate dehydrogenase complex is activated to funnel glucose to pyruvate and then to acetyl CoA and ultimately to fatty acids.

In people with a phosphatase deficiency, pyruvate dehydrogenase is always phosphorylated and thus inactive. Consequently, glucose is processed to lactate rather than acetyl CoA. This condition results in unremitting lactic acidosis—high blood levels of lactic acid. In such an acidic environment, many tissues malfunction, most notably the central nervous system (Problems 13 and 14).

The citric acid cycle is controlled at several points

The rate of the citric acid cycle is precisely adjusted to meet an animal cell's needs for ATP (Figure 17.19). The primary control points are the allosteric enzymes isocitrate dehydrogenase and α-ketoglutarate dehydrogenase, the first two enzymes in the cycle to generate high-energy electrons.

The first control site is isocitrate dehydrogenase. The enzyme is allosterically stimulated by ADP, which enhances the enzyme's affinity for substrates. The binding of isocitrate, NAD^+, Mg^{2+}, and ADP is mutually cooperative. In contrast, ATP is inhibitory. The reaction product NADH also inhibits isocitrate dehydrogenase by directly displacing NAD^+. It is important to note that several steps in the cycle require NAD^+ or FAD, which are abundant only when the energy charge is low.

A second control site in the citric acid cycle is α-ketoglutarate dehydrogenase, which catalyzes the rate-limiting step in the citric acid cycle. Some aspects of this enzyme's control are like those of the pyruvate dehydrogenase complex, as might be expected from the homology of the two enzymes. α-Ketoglutarate dehydrogenase is inhibited by succinyl CoA and NADH, the products of the reaction that it catalyzes. In addition, α-ketoglutarate dehydrogenase is inhibited by a high energy charge. Thus, *the rate of the cycle is reduced when the cell has a high level of ATP.* α-Ketoglutarate dehydrogenase deficiency is observed in a number of neurological disorders, including Alzheimer disease.

The use of isocitrate dehydrogenase and α-ketoglutarate dehydrogenase as control points integrates the citric acid cycle with other pathways and highlights the central role of the citric acid cycle in metabolism. For instance, the inhibition of isocitrate dehydrogenase leads to a buildup of citrate, because the interconversion of isocitrate and citrate is readily reversible under intracellular conditions. Citrate can be transported to the cytoplasm, where it signals phosphofructokinase to halt glycolysis (Section 16.2) and where it can serve as a source of acetyl CoA for fatty acid synthesis (Section 22.4). The α-ketoglutarate that accumulates when α-ketoglutarate

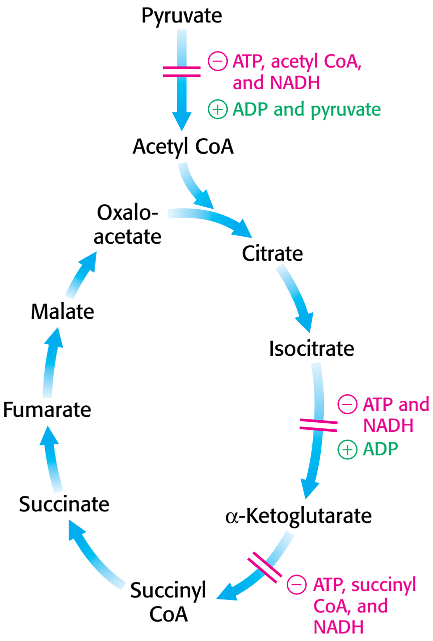


FIGURE 17.19 Control of the citric acid cycle. The citric acid cycle is regulated primarily by the concentration of ATP and NADH. The key control points are the enzymes isocitrate dehydrogenase and α-ketoglutarate dehydrogenase.

dehydrogenase is inhibited can be used as a precursor for several amino acids and the purine bases (Chapter 23 and Chapter 25).

In many bacteria, the funneling of two-carbon fragments into the cycle also is controlled. *The synthesis of citrate from oxaloacetate and acetyl CoA carbon units is an important control point in these organisms.* ATP is an allosteric inhibitor of citrate synthase. The effect of ATP is to increase the value of K_M for acetyl CoA. Thus, as the level of ATP increases, less of this enzyme is bound to acetyl CoA and so less citrate is formed.

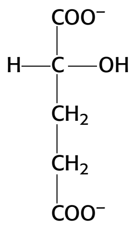
Defects in the citric acid cycle contribute to the development of cancer

Four enzymes crucial to cellular respiration are known to contribute to the development of cancer: succinate dehydrogenase, fumarase, pyruvate dehydrogenase kinase and isocitrate dehydrogenase. Mutations that alter the activity of the first three of these enzymes enhance aerobic glycolysis (Section 16.2). In aerobic glycolysis, cancer cells preferentially metabolize glucose to lactate even in the presence of oxygen. Defects in these enzymes share a common biochemical link: the transcription factor *hypoxia inducible factor 1* (HIF-1).

Normally, HIF-1 up-regulates the enzymes and transporters that enhance glycolysis only when oxygen concentration falls, a condition called hypoxia. Under normal conditions, HIF-1 is hydroxylated by prolyl hydroxylase 2 and is subsequently destroyed by the proteasome, a large complex of proteolytic enzymes (Chapter 23). The degradation of HIF-1 prevents the stimulation of glycolysis. Prolyl hydroxylase 2 requires α-ketoglutarate, ascorbate, and oxygen for activity. Thus, when oxygen concentration falls, the prolyl hydroxylase 2 is inactive, HIF-1 is not hydroxylated and not degraded, and the synthesis of proteins required for glycolysis is stimulated. As a result, the rate of glycolysis is increased.

Recent research suggestions that defects in the enzymes of the citric acid cycle may significantly affect the regulation of prolyl hydroxylase 2. When either succinate dehydrogenase or fumarase is defective, succinate and fumarate accumulate in the mitochondria and spill over into the cytoplasm. Both succinate and fumarate are competitive inhibitors of prolyl hydroxylase 2. The inhibition of prolyl hydroxylase 2 results in the stabilization of HIF-1, since HIF-1 is no longer hydroxylated. Lactate, the end product of glycolysis, also appears to inhibit prolyl hydroxylase 2 by interfering with the action of ascorbate. In addition to increasing the amount of the proteins required for glycolysis, HIF-1 also stimulates the production of pyruvate dehydrogenase kinase (PDK). The kinase inhibits the pyruvate dehydrogenase complex, preventing the conversion of pyruvate into acetyl CoA. The pyruvate remains in the cytoplasm, further increasing the rate of aerobic glycolysis. Moreover, mutations in PDK that lead to enhanced activity contribute to increased aerobic glycolysis and the subsequent development of cancer. By enhancing glycolysis and increasing the concentration of lactate, the mutations in PDK result in the inhibition of hydroxylase and the stabilization of HIF-1.

Mutations in isocitrate dehydrogenase result in the generation of an oncogenic metabolite, 2-hydroxyglutarate. The mutant enzyme catalyzes the conversion of isocitrate to α-ketoglutarate, but then reduces α-ketoglutarate to form 2-hydroxyglutarate. 2-Hydroxyglutarate alters the methylation patterns in DNA (Section 32.2) and reduces dependence on growth factors for proliferation. These changes alter gene expression and promote unrestrained cell growth.



2-Hydroxyglutarate

These observations linking citric acid cycle enzymes to cancer suggest that cancer is also a metabolic disease, not simply a disease of mutant growth factors and cell cycle control proteins. The realization that there is a metabolic component to cancer opens the door to new thinking about the control of cancer. Indeed, preliminary experiments suggest that if cancer cells undergoing aerobic glycolysis are forced by pharmacological manipulation to use oxidative phosphorylation, the cancer cells lose their malignant properties. It is also interesting to note that the citric acid cycle, which has been studied for decades, still has secrets to be revealed by future biochemists.

17.4 The Citric Acid Cycle Is a Source of Biosynthetic Precursors

Thus far, discussion has focused on the citric acid cycle as the *major degradative pathway for the generation of ATP.* As a major metabolic hub of the cell, the citric acid cycle also *provides intermediates for biosyntheses* (Figure 17.20). For example, most of the carbon atoms in porphyrins come from *succinyl CoA.* Many of the amino acids are derived from *α-ketoglutarate* and *oxaloacetate.* These biosynthetic processes will be considered in subsequent chapters.

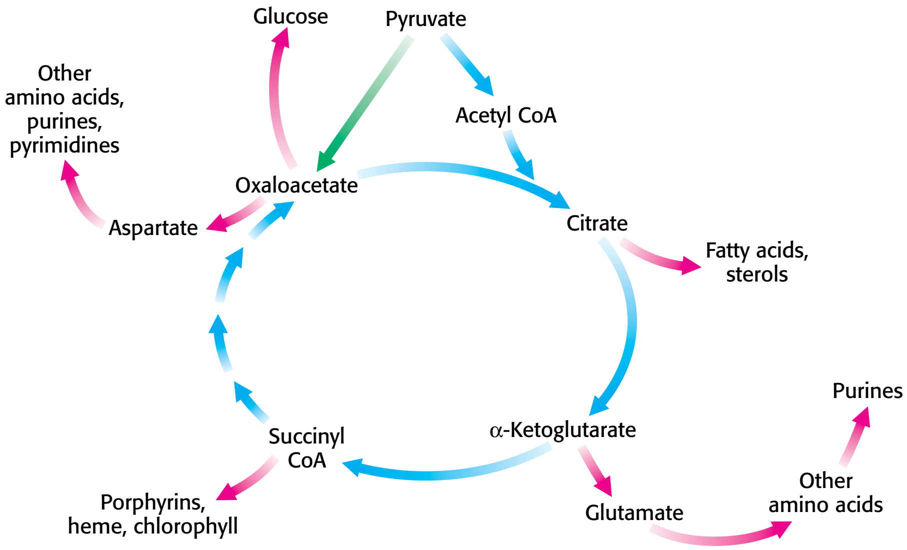

Glucose
Pyruvate
Other amino acids, purines, pyrimidines
Acetyl CoA
Oxaloacetate
Aspartate
Citrate
Fatty acids, sterols
Purines
Succinyl CoA
α-Ketoglutarate
Other amino acids
Porphyrins, heme, chlorophyll
Glutamate

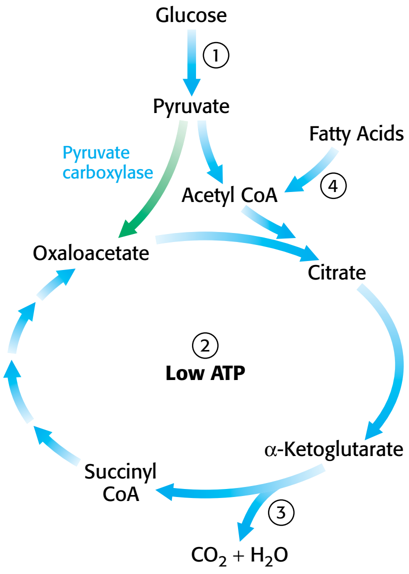
FIGURE 17.20 Biosynthetic roles of the citric acid cycle. Intermediates are drawn off for biosyntheses (shown by red arrows) when the energy needs of the cell are met. Intermediates are replenished by the formation of oxaloacetate from pyruvate.

The citric acid cycle must be capable of being rapidly replenished

An important consideration is that *citric acid cycle intermediates must be replenished if any are drawn off for biosyntheses.* Suppose that much oxaloacetate is converted into amino acids for protein synthesis and, subsequently, the energy needs of the cell rise. The citric acid cycle will operate to a reduced extent unless new oxaloacetate is formed because acetyl CoA cannot enter the cycle unless it condenses with oxaloacetate. Even though oxaloacetate is recycled, a minimal level must be maintained to allow the cycle to function.

How is oxaloacetate replenished? Mammals lack the enzymes for the net conversion of acetyl CoA into oxaloacetate or any other citric acid cycle intermediate. Rather, oxaloacetate is formed by the carboxylation of pyruvate, in a reaction catalyzed by the biotin-dependent enzyme *pyruvate carboxylase* (Figure 17.21).

$$\text{Pyruvate} + \text{CO}_2 + \text{ATP} + \text{H}_2\text{O} \rightarrow \text{oxaloacetate} + \text{ADP} + \text{P}_i + 2\text{H}^+$$


Glucose
①
Pyruvate
Fatty Acids
Pyruvate carboxylase
Acetyl CoA
④
Oxaloacetate
Citrate
②
Low ATP
α-Ketoglutarate
Succinyl CoA
③
CO2 + H2O

Active pathways

① Glycolysis, Ch. 16
② Citric acid cycle, Ch. 17
③ Oxidative phosphorylation, Ch. 18
④ Fatty acid oxidation, Ch. 22

FIGURE 17.21 PATHWAY INTEGRATION: Pathways active during exercise after a night's rest. The rate of the citric acid cycle increases during exercise, requiring the replenishment of oxaloacetate and acetyl CoA. Oxaloacetate is replenished by its formation from pyruvate. Acetyl CoA may be produced from the metabolism of both pyruvate and fatty acids.

Recall that this enzyme plays a crucial role in gluconeogenesis (Section 16.3). It is active only in the presence of acetyl CoA, which signifies the need for more oxaloacetate. If the energy charge is high, oxaloacetate is converted into glucose. If the energy charge is low, oxaloacetate replenishes the citric acid cycle. The synthesis of oxaloacetate by the carboxylation of pyruvate is an example of an *anaplerotic reaction* (*anaplerotic* is of Greek origin, meaning to "fill up"), a reaction that leads to the net synthesis, or replenishment, of pathway components. Note that because the citric acid cycle is a cycle, it can be replenished by the generation of any of the intermediates. Glutamine is an especially important source of citric acid cycle intermediates in rapidly growing cells, including cancer cells. Glutamine is converted into glutamate and then into α-ketoglutarate.

The disruption of pyruvate metabolism is the cause of beriberi and poisoning by mercury and arsenic

Beriberi, a neurologic and cardiovascular disorder, is caused by a dietary deficiency of thiamine (also called *vitamin B_1*). The disease has been and continues to be a serious health problem in the Far East because rice, the major food, has a rather low content of thiamine. This deficiency is partly ameliorated if the whole rice grain is soaked in water before milling; some of the thiamine in the husk then leaches into the rice kernel. The problem is exacerbated if the rice is polished to remove the outer layer (that is, converted from brown to white rice), because only the outer layer contains significant amounts of thiamine. A form of beriberi, called Wernicke's encephalopathy, is also occasionally seen in alcoholics who are severely malnourished and thus thiamine deficient. The disease is characterized by neurologic and cardiac symptoms. Damage to the peripheral nervous system is expressed as pain in the limbs, weakness of the musculature, and distorted skin sensation. The heart may be enlarged and the cardiac output inadequate.

Beriberi

A vitamin-deficiency disease first described in 1630 by Jacob Bonitus, a Dutch physician working in Java:

"A certain very troublesome affliction, which attacks men, is called by the inhabitants Beriberi (which means sheep). I believe those, whom this same disease attacks, with their knees shaking and the legs raised up, walk like sheep. It is a kind of paralysis, or rather Tremor: for it penetrates the motion and sensation of the hands and feet indeed sometimes of the whole body."

Which biochemical processes might be affected by a deficiency of thiamine? Thiamine is the precursor of the cofactor thiamine pyrophosphate. *This cofactor is the prosthetic group of three important enzymes: pyruvate dehydrogenase, α-ketoglutarate dehydrogenase, and transketolase.* Transketolase functions in the pentose phosphate pathway, which will be considered in Chapter 20. *The common feature of enzymatic reactions utilizing TPP is the transfer of an activated aldehyde unit.* In beriberi, the levels of pyruvate and α-ketoglutarate in the blood are higher than normal. The increase in the level of pyruvate in the blood is especially pronounced after the ingestion of glucose. A related finding is that the activities of the pyruvate and α-ketoglutarate dehydrogenase complexes in vivo are abnormally low. The low transketolase activity of red blood cells in beriberi is an easily measured and reliable diagnostic indicator of the disease.

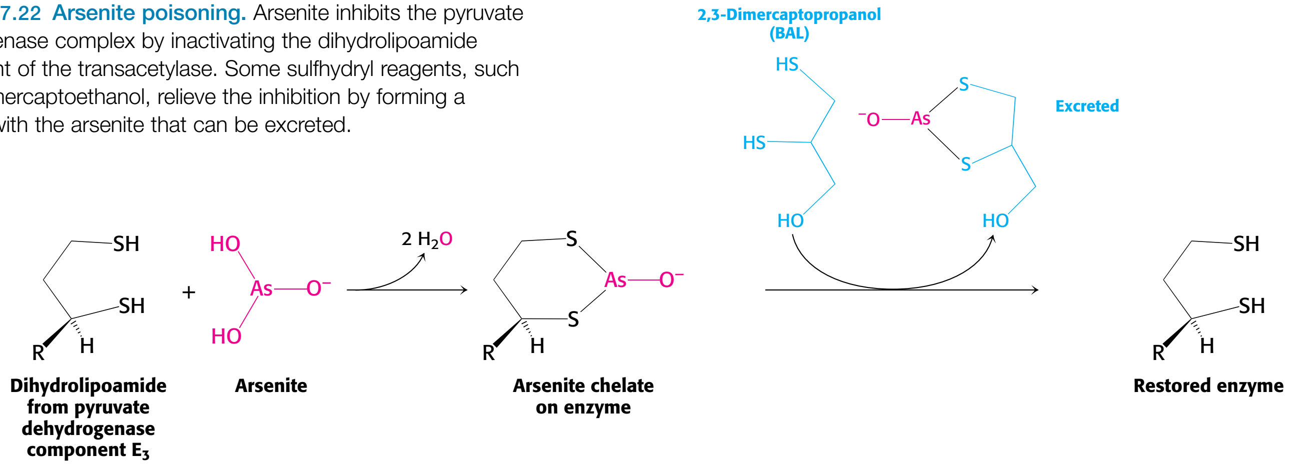
Why does TPP deficiency lead primarily to neurological disorders? The nervous system relies essentially on glucose as its only fuel. The product of glycolysis, pyruvate, can enter the citric acid cycle only through the pyruvate dehydrogenase complex. With that enzyme deactivated, the nervous system has no source of fuel. In contrast, most other tissues can use fats as a source of fuel for the citric acid cycle.

Symptoms similar to those of beriberi appear in organisms exposed to mercury or arsenite (AsO_3^{3-}). Both materials have a high affinity for neighboring sulfhydryls, such as those in the reduced dihydrolipoyl groups of the E_3 component of the pyruvate dehydrogenase complex (Figure 17.22). The binding of mercury or arsenite to the dihydrolipoyl groups inhibits the complex and leads to central nervous system pathologies. The proverbial phrase "mad as a hatter" refers to the strange behavior of poisoned hat makers who

In this Style 10/6

[The Granger Collection.]

FIGURE 17.22 Arsenite poisoning. Arsenite inhibits the pyruvate dehydrogenase complex by inactivating the dihydrolipoamide component of the transacetylase. Some sulfhydryl reagents, such as 2,3-dimercaptoethanol, relieve the inhibition by forming a complex with the arsenite that can be excreted.


2,3-Dimercaptopropanol (BAL)
HS
HS
HO
⁻O
As
S
S
Excreted
HO
SH
SH
R
H
+
HO
As
O⁻
HO
2 H_2O
S
As
O⁻
S
R
H
SH
SH
R
H
Dihydrolipoamide from pyruvate dehydrogenase component E_3
Arsenite
Arsenite chelate on enzyme
Restored enzyme

used mercury nitrate to soften and shape animal furs. This form of mercury is absorbed through the skin. Similar symptoms afflicted the early photographers, who used vaporized mercury to create daguerreotypes.

Treatment for these poisons is the administration of sulfhydryl reagents with adjacent sulfhydryl groups to compete with the dihydrolipoyl residues for binding with the metal ion. The reagent–metal complex is then excreted in the urine. Indeed, 2,3-dimercaptopropanol (Figure 17.22) was developed after World War I as an antidote to lewisite, an arsenic-based chemical weapon. This compound was initially called BAL, for British anti-lewisite.

The citric acid cycle may have evolved from preexisting pathways

How did the citric acid cycle come into being? Although definitive answers are elusive, informed speculation is possible. We can perhaps begin to comprehend how evolution might work at the level of biochemical pathways.

The citric acid cycle was most likely assembled from preexisting reaction pathways. As noted earlier, many of the intermediates formed in the citric acid cycle are used in metabolic pathways for amino acids and porphyrins. Thus, compounds such as pyruvate, α-ketoglutarate, and oxaloacetate were likely present early in evolution for biosynthetic purposes. The oxidative decarboxylation of these α-ketoacids is quite favorable thermodynamically and can be used to drive the synthesis of both acyl CoA derivatives and NADH. These reactions almost certainly formed the core of processes that preceded the citric acid cycle evolutionarily. Interestingly, α-ketoglutarate and oxaloacetate can be interconverted by transamination of the respective amino acids by aspartate aminotransferase, another key biosynthetic enzyme. Thus, cycles comprising smaller numbers of intermediates used for a variety of biochemical purposes could have existed before the present form evolved.

The manuscript proposing the citric acid cycle was submitted for publication to *Nature* but was rejected in June 1937. That same year it was published in *Enzymologia.* Dr. Krebs proudly displayed the rejection letter throughout his career as encouragement for young scientists.

"The editor of NATURE presents his compliments to Dr. H. A. Krebs and regrets that as he has already sufficient letters to fill the correspondence columns of NATURE for seven or eight weeks, it is undesirable to accept further letters at the present time on account of the time delay which must occur in their publication.

If Dr. Krebs does not mind much delay the editor is prepared to keep the letter until the congestion is relieved in the hope of making use of it.

He returns it now, in case Dr. Krebs prefers to submit it for early publication to another periodical."

17.5 The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate

Acetyl CoA that enters the citric acid cycle has but one fate: oxidation to CO_2 and H_2O. Most organisms thus cannot convert acetyl CoA into glucose. Although oxaloacetate, a key precursor to glucose, is formed in the citric acid cycle, the two decarboxylations that take place before the regeneration of oxaloacetate preclude the *net* conversion of acetyl CoA into glucose.

In plants and in some microorganisms, there is a metabolic pathway that allows the conversion of acetyl CoA generated from fats stores into glucose. This reaction sequence, called the *glyoxylate cycle,* is similar to the citric acid cycle but bypasses the two decarboxylation steps of the cycle. Another important difference is that two molecules of acetyl CoA enter per turn of the glyoxylate cycle, compared with one in the citric acid cycle.

The glyoxylate cycle (Figure 17.23), like the citric acid cycle, begins with the condensation of acetyl CoA and oxaloacetate to form citrate, which is then isomerized to isocitrate. Instead of being decarboxylated, as in the citric acid cycle, isocitrate is cleaved by *isocitrate lyase* into succinate and glyoxylate. The ensuing steps regenerate oxaloacetate from glyoxylate. First, acetyl CoA condenses with glyoxylate to form malate in a reaction catalyzed by *malate synthase,* and then malate is oxidized to oxaloacetate, as in the citric acid cycle. The sum of these reactions is

$$2\text{ Acetyl CoA} + \text{NAD}^+ + 2\,\text{H}_2\text{O} \longrightarrow \text{succinate} + 2\text{ CoA} + \text{NADH} + 2\,\text{H}^+$$

In plants, these reactions take place in organelles called *glyoxysomes.* This cycle is especially prominent in oil-rich seeds, such as those from sunflowers, cucumbers, and castor beans. Succinate, released midcycle, can be converted into carbohydrates by a combination of the citric acid cycle and gluconeogenesis. The carbohydrates power seedling growth until the cell can begin photosynthesis. Thus, organisms with the glyoxylate cycle gain a metabolic versatility because they can use acetyl CoA as a precursor of glucose and other biomolecules.

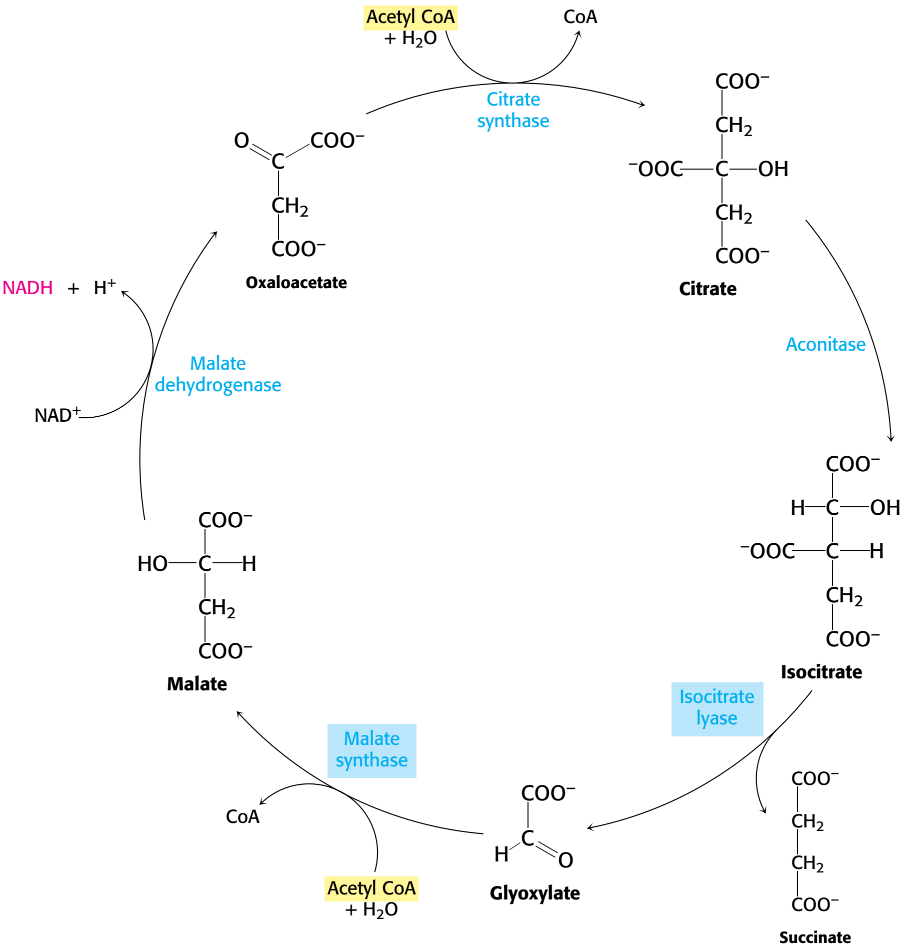


FIGURE 17.23 The glyoxylate cycle. The glyoxylate cycle allows plants and some microorganisms to grow on acetate because the cycle bypasses the decarboxylation steps of the citric acid cycle. The reactions of this cycle are the same as those of the citric acid cycle except for the ones catalyzed by isocitrate lyase and malate synthase, which are boxed in blue.

SUMMARY

The citric acid cycle is the final common pathway for the oxidation of fuel molecules. It also serves as a source of building blocks for biosyntheses.

17.1 The Pyruvate Dehydrogenase Complex Links Glycolysis to the Citric Acid Cycle

Most fuel molecules enter the cycle as acetyl CoA. The link between glycolysis and the citric acid cycle is the oxidative decarboxylation of pyruvate to form acetyl CoA. In eukaryotes, this reaction and those of the cycle take place inside mitochondria, in contrast with glycolysis, which takes place in the cytoplasm.

17.2 The Citric Acid Cycle Oxidizes Two-Carbon Units

The cycle starts with the condensation of oxaloacetate (C_4) and the acetyl unit (C_2) of acetyl CoA to give citrate (C_6), which is isomerized to isocitrate (C_6). Oxidative decarboxylation of this intermediate gives α-ketoglutarate (C_5). The second molecule of carbon dioxide comes off in the next reaction, in which α-ketoglutarate is oxidatively decarboxylated to succinyl CoA (C_4). The thioester bond of succinyl CoA is cleaved by orthophosphate to yield succinate, and an ATP is concomitantly generated. Succinate is oxidized to fumarate (C_4), which is then hydrated to form malate (C_4). Finally, malate is oxidized to regenerate oxaloacetate (C_4). Thus, two carbon atoms from acetyl CoA enter the cycle, and two carbon atoms leave the cycle as CO_2 in the successive decarboxylations catalyzed by isocitrate dehydrogenase and α-ketoglutarate dehydrogenase. In the four oxidation–reduction reactions in the cycle, three pairs of electrons are transferred to NAD^+ and one pair to FAD. These reduced electron carriers are subsequently oxidized by the electron-transport chain to generate approximately 9 molecules of ATP. In addition, 1 molecule of ATP is directly formed in the citric acid cycle. Hence, a total of 10 molecules of ATP are generated for each two-carbon fragment that is completely oxidized to H_2O and CO_2.

17.3 Entry to the Citric Acid Cycle and Metabolism Through It Are Controlled

The citric acid cycle operates only under aerobic conditions because it requires a supply of NAD^+ and FAD. These electron acceptors are regenerated when NADH and $FADH_2$ transfer their electrons to O_2 through the electron-transport chain, with the concomitant production of ATP. Consequently, the rate of the citric acid cycle depends on the need for ATP. The irreversible formation of acetyl CoA from pyruvate is an important regulatory point for the entry of glucose-derived pyruvate into the citric acid cycle. The activity of the pyruvate dehydrogenase complex is stringently controlled by reversible phosphorylation. In eukaryotes, the regulation of two enzymes in the cycle also is important for control. A high energy charge diminishes the activities of isocitrate dehydrogenase and α-ketoglutarate dehydrogenase. These mechanisms complement each other in reducing the rate of formation of acetyl CoA when the energy charge of the cell is high and when biosynthetic intermediates are abundant.

17.4 The Citric Acid Cycle Is a Source of Biosynthetic Precursors

When the cell has adequate energy available, the citric acid cycle can also provide a source of building blocks for a host of important biomolecules, such as nucleotide bases, proteins, and heme groups. This use depletes the cycle of intermediates. When the cycle again needs to metabolize fuel, anaplerotic reactions replenish the cycle intermediates.

17.5 The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate

The glyoxylate cycle enhances the metabolic versatility of many plants and bacteria. This cycle, which uses some of the reactions of the citric acid cycle, enables these organisms to subsist on acetate because it bypasses the two decarboxylation steps of the citric acid cycle.

KEY TERMS

citric acid (tricarboxylic acid, TCA; Krebs) cycle (p. 495)
acetyl CoA (p. 495)
oxidative phosphorylation (p. 497)
pyruvate dehydrogenase complex (p. 497)
flavoprotein (p. 500)
citrate synthase (p. 502)
iron–sulfur (nonheme iron) protein (p. 504)
isocitrate dehydrogenase (p. 504)
α-ketoglutarate dehydrogenase (p. 505)
anaplerotic reaction (p. 515)
beriberi (p. 515)
glyoxylate cycle (p. 517)
isocitrate lyase (p. 517)
malate synthase (p. 517)
glyoxysome (p. 517)

PROBLEMS

1. *A one-way link.* What reaction serves to link glycolysis and the citric acid cycle, and what is the enzyme that catalyzes the reaction?

2. *Naming names.* What are the five enzymes (including regulatory enzymes) that constitute the pyruvate dehydrogenase complex? Which reactions do they catalyze?

3. *Coenzymes.* What coenzymes are required by the pyruvate dehydrogenase complex? What are their roles?

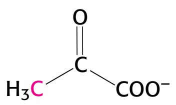
4. *More coenzymes.* Distinguish between catalytic coenzymes and stoichiometric coenzymes in the pyruvate dehydrogenase complex.

5. *Waste and fraud?* Figure 17.9 shows the steps in the pyruvate dehydrogenase complex reaction cycle. A key product, acetyl CoA, is released after the fourth step. What is the purpose of the remaining steps?

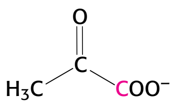
6. *Joined at the hip.* List some of the advantages of organizing the enzymes that catalyze the formation of acetyl CoA from pyruvate into a single large complex.

7. *Flow of carbon atoms.* What is the fate of the radioactive label when each of the following compounds is added to a cell extract containing the enzymes and cofactors of the glycolytic pathway, the citric acid cycle, and the pyruvate dehydrogenase complex? (The ^{14}C label is printed in red.)

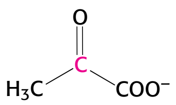
(a)



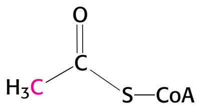
(b)



(c)



(d)



(e) Glucose 6-phosphate labeled at C-1.

8. $C_2 + C_2 \rightarrow C_4$

(a) Which enzymes are required to get *net synthesis* of oxaloacetate from acetyl CoA?

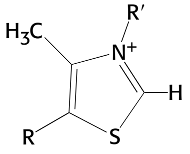
(b) Write a balanced equation for the net synthesis.

(c) Do mammalian cells contain the requisite enzymes?

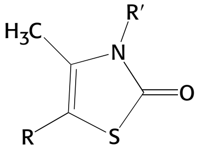
9. *Driving force.* What is the $\Delta G^{\circ\prime}$ for the complete oxidation of the acetyl unit of acetyl CoA by the citric acid cycle?

10. *Acting catalytically.* The citric acid cycle itself, which is composed of enzymatically catalyzed steps, can be thought of essentially as a supramolecular enzyme. Explain.

11. *A potent inhibitor.* Thiamine thiazolone pyrophosphate binds to pyruvate dehydrogenase about 20,000 times as strongly as does thiamine pyrophosphate, and it competitively inhibits the enzyme. Why?



TPP



Thiazolone analog of TPP

12. *Like Holmes and Watson.* Match each term with its description.

(a) Acetyl CoA _____
(b) Citric acid cycle _____
(c) Pyruvate dehydrogenase complex _____
(d) Thiamine pyrophosphate _____
(e) Lipoic acid _____
(f) Pyruvate dehydrogenase _____
(g) Acetyllipoamide _____
(h) Dihydrolipoyl transacetylase _____
(i) Dihydrolipoyl dehydrogenase _____
(j) Beriberi _____

1. Catalyzes the link between glycolysis and the citric acid cycle.
2. Coenzyme required by transacetylase
3. Final product of pyruvate dehydrogenase
4. Catalyzes the formation of acetyl CoA
5. Regenerates active transacetylase
6. Fuel for the citric acid cycle
7. Coenzyme required by pyruvate dehydrogenase
8. Catalyzes the oxidative decarboxylation of pyruvate
9. Due to a deficiency of thiamine
10. Central metabolic hub

13. *Lactic acidosis.* Patients in shock often suffer from lactic acidosis owing to a deficiency of O_2. Why does a lack of O_2 lead to lactic acid accumulation? One treatment for shock is to administer dichloroacetate (DCA), which inhibits the kinase associated with the pyruvate dehydrogenase complex. What is the biochemical rationale for this treatment?

14. *DCA again.* Patients with pyruvate dehydrogenase deficiency show high levels of lactic acid in the blood. However, in some cases, treatment with dichloroacetate (DCA) lowers lactic acid levels.

(a) How does DCA act to stimulate pyruvate dehydrogenase activity?

(b) What does this suggest about pyruvate dehydrogenase activity in patients who respond to DCA?

15. *Energy rich.* What are the thioesters in the reaction catalyzed by PDH complex?

16. *Alternative fates.* Compare the regulation of the pyruvate dehydrogenase complex in muscle and in liver.

17. *Mutations.*

(a) Predict the effect of a mutation that enhances the activity of the kinase associated with the PDH complex.

(b) Predict the effect of a mutation that reduces the activity of the phosphatase associated with the PDH complex.

18. *Flaking paint, green wallpaper.* Clare Boothe Luce, ambassador to Italy in the 1950s (and Connecticut congressperson, playwright, editor of *Vanity Fair,* and the wife of Henry Luce, founder of *Time* magazine and *Sports Illustrated*) became ill when she was staying at the ambassadorial residence in Rome. The paint on the dining room ceiling, an arsenic-based paint, was flaking; the wallpaper of her bedroom in the ambassadorial residence was colored a mellow green owing to the presence of cupric arsenite in the pigment. Suggest a possible cause of Ambassador Luce's illness.

19. *Like Jack and Jill.* Match each enzyme with its description.

(a) Pyruvate dehydrogenase complex _____
(b) Citrate synthase _____
(c) Aconitase _____
(d) Isocitrate dehydrogenase _____
(e) α-Ketoglutarate dehydrogenase _____
(f) Succinyl CoA synthetase _____
(g) Succinate dehydrogenase _____
(h) Fumarase _____
(i) Malate dehydrogenase _____
(j) Pyruvate carboxylase _____

1. Catalyzes the formation of isocitrate
2. Synthesizes succinyl CoA
3. Generates malate
4. Generates ATP
5. Converts pyruvate into acetyl CoA
6. Converts pyruvate into oxaloacetate
7. Condenses oxaloacetate and acetyl CoA
8. Catalyzes the formation of oxaloacetate
9. Synthesizes fumarate
10. Catalyzes the formation of α-ketoglutarate

20. *A hoax, perhaps?* The citric acid cycle is part of aerobic respiration, but no O_2 is required for the cycle. Explain this paradox.

21. *One of a kind.* How is succinate dehydrogenase unique compared with the other enzymes in the citric acid cycle?

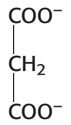
22. *Coupling reactions.* The oxidation of malate by NAD^+ to form oxaloacetate is a highly endergonic reaction under standard conditions [$\Delta G^{\circ\prime} = 29\ \text{kJ mol}^{-1}\ (7\ \text{kcal mol}^{-1})$]. The reaction proceeds readily under physiological conditions.

(a) Explain why the reaction proceeds as written under physiological conditions.

(b) Assuming an $[NAD^+]/[NADH]$ ratio of 8 and a pH of 7, what is the lowest [malate]/[oxaloacetate] ratio at which oxaloacetate can be formed from malate?

23. *Synthesizing α-ketoglutarate.* It is possible, with the use of the reactions and enzymes considered in this chapter, to convert pyruvate into α-ketoglutarate without depleting any of the citric acid cycle components. Write a balanced reaction scheme for this conversion, showing cofactors and identifying the required enzymes.

24. *Seven o'clock roadblock.* Malonate is a competitive inhibitor of succinate dehydrogenase. How will the concentrations of citric acid cycle intermediates change immediately after the addition of malonate? Why is malonate not a substrate for succinate dehydrogenase?



$$\mathrm{COO^-}\text{—}\mathrm{CH_2}\text{—}\mathrm{COO^-}$$

Malonate

25. *No signal, no activity.* Why is acetyl CoA an especially appropriate activator for pyruvate carboxylase?

26. *Power differentials.* As we will see in the next chapter, when NADH reacts with oxygen, 2.5 ATP are generated. When $FADH_2$ reduces oxygen, only 1.5 ATP are generated. Why then does succinate dehydrogenase produce $FADH_2$ and not NADH when succinate is reduced to fumarate?

27. *Back to Orgo (or do you say O chem?).* Before any oxidation can occur in the citric acid cycle, citrate must be isomerized into isocitrate. Why is this the case?

28. *A nod is as good as a wink to a blind horse.* Explain why a GTP molecule, or another nucleoside triphosphate, is energetically equivalent to an ATP molecule in metabolism.

29. *One from two.* The synthesis of citrate from acetyl CoA and oxaloacetate is a biosynthetic reaction. What is the energy source that drives formation of citrate?

30. *Versatility.* What is the chief benefit of being able to perform the glyoxylate cycle?

Chapter Integration Problems

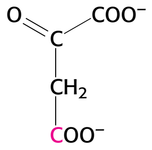
31. *Fats into glucose?* Fats are usually metabolized into acetyl CoA and then further processed through the citric acid cycle. In Chapter 16, we saw that glucose can be synthesized from oxaloacetate, a citric acid cycle intermediate. Why, then, after a long bout of exercise depletes our carbohydrate stores, do we need to replenish those stores by eating carbohydrates? Why do we not simply replace them by converting fats into carbohydrates?

32. *Alternative fuels*. As we will see (Chapter 22), fatty acid breakdown generates a large amount of acetyl CoA. What will be the effect of fatty acid breakdown on pyruvate dehydrogenase complex activity? On glycolysis?

Mechanism Problems

33. *Theme and variation*. Propose a reaction mechanism for the condensation of acetyl CoA and glyoxylate in the glyoxylate cycle of plants and bacteria.

34. *Symmetry problems*. In experiments carried out in 1941 to investigate the citric acid cycle, oxaloacetate labeled with ^{14}C in the carboxyl carbon atom farthest from the keto group was introduced to an active preparation of mitochondria.



Oxaloacetate

Analysis of the α-ketoglutarate formed showed that none of the radioactive label had been lost. Decarboxylation of α-ketoglutarate then yielded succinate devoid of radioactivity. All the label was in the released CO_2. Why were the early investigators of the citric acid cycle surprised that *all* the label emerged in the CO_2?

35. *Symmetric molecules reacting asymmetrically.* The interpretation of the experiments described in Problem 34 was that citrate (or any other symmetric compound) cannot be an intermediate in the formation of α-ketoglutarate, because of the asymmetric fate of the label. This view seemed compelling until Alexander Ogston incisively pointed out in 1948 that "it is possible that *an asymmetric enzyme which attacks a symmetrical compound can distinguish between its identical groups* [italics added]." For simplicity, consider a molecule in which two hydrogen atoms, a group X, and a different group Y are bonded to a tetrahedral carbon atom as a model for citrate. Explain how a symmetric molecule can react with an enzyme in an asymmetric way.

Data Interpretation Problem

36. *A little goes a long way.* As will become clearer in Chapter 18, the activity of the citric acid cycle can be monitored by measuring the amount of O_2 consumed. The greater the rate of O_2 consumption, the faster the rate of the cycle. Hans Krebs used this assay to investigate the cycle in 1937. He used as his experimental system minced pigeon-breast muscle, which is rich in mitochondria. In one set of experiments, Krebs measured the O_2 consumption in the presence of carbohydrate only and in the presence of carbohydrate and citrate. The results are shown in the following table.

Effect of citrate on oxygen consumption by minced pigeon-breast muscle

	Micromoles of oxygen consumed	
Time (min)	Carbohydrate only	Carbohydrate plus 3 µmol of citrate
10	26	28
60	43	62
90	46	77
150	49	85

(a) How much O_2 would be absorbed if the added citrate were completely oxidized to H_2O and CO_2?

(b) On the basis of your answer to part *a*, what do the results given in the table suggest?

37. *Arsenite poisoning*. The effect of arsenite on the experimental system of Problem 36 was then examined. Experimental data (not presented here) showed that the amount of citrate present did not change in the course of the experiment in the absence of arsenite. However, if arsenite was added to the system, different results were obtained, as shown in the following table.

Disappearance of citric acid in pigeon-breast muscle in the presence of arsenite

Micromoles of citrate added	Micromoles of citrate found after 40 minutes	Micromoles of citrate used
22	0.6	21
44	20.0	24
90	56.0	34

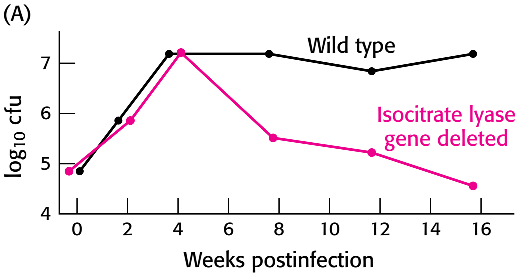
(a) What is the effect of arsenite on the disappearance of citrate?

(b) How is the action of arsenite altered by the addition of more citrate?

(c) What do these data suggest about the site of action of arsenite?

38. *Isocitrate lyase and tuberculosis.* The bacterium *Mycobacterium tuberculosis,* the cause of tuberculosis, can invade the lungs and persist in a latent state for years. During this time, the bacteria reside in granulomas—nodular scars containing bacteria and host-cell debris in the center and surrounded by immune cells. The granulomas are lipid-rich, oxygen-poor environments. How these bacteria manage to persist is something of a mystery. The results of recent research suggest that the glyoxylate cycle is required for the persistence. The following data show the amount of bacteria

[presented as colony-forming units (cfu)] in mice lungs in the weeks after an infection.



In graph A, the black circles represent the results for wild-type bacteria and the red circles represent the results for bacteria from which the gene for isocitrate lyase was deleted.

(a) What is the effect of the absence of isocitrate lyase?

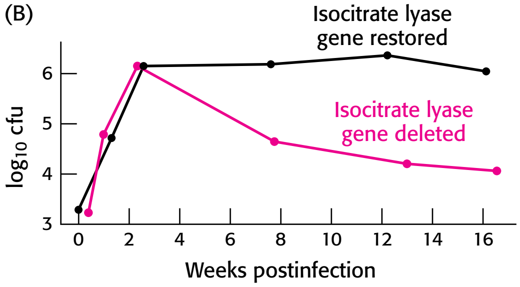
The techniques described in Chapter 5 were used to reinsert the gene encoding isocitrate lyase into bacteria from which it had previously been deleted.

In graph B, black circles represent bacteria into which the gene was reinserted and red circles represent bacteria in which the gene was still missing.

(b) Do these results support those obtained in part *a*?

(c) What is the purpose of the experiment in part *b*?

(d) Why do these bacteria perish in the absence of the glyoxylate cycle?

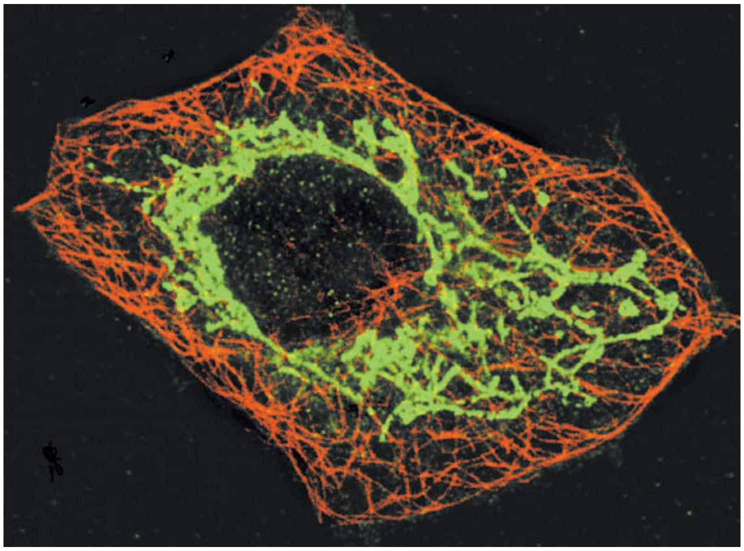


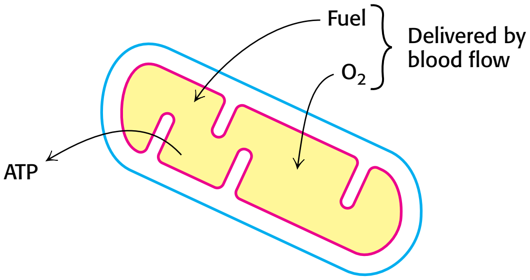
[Data from KcKinney et al., *Nature* 406:735–738, 2000.]

CHAPTER

18

Oxidative Phosphorylation





Mitochondria, stained green, form a network inside a fibroblast cell (left). Mitochondria oxidize carbon fuels to form cellular energy in the form of ATP. [(Left) Courtesy of Michael P. Yaffee, Department of Biology, University of California at San Diego.]

The amount of ATP that human beings require to go about their lives is staggering. A sedentary male of 70 kg (154 lbs) requires about 8400 kJ (2000 kcal) for a day's worth of activity. To provide this much energy requires 83 kg of ATP. However, human beings possess only about 250 g of ATP at any given moment. The disparity between the amount of ATP that we have and the amount that we require is compensated by recycling ADP back to ATP. Each ATP molecule is recycled approximately 300 times per day. This recycling takes place primarily through *oxidative phosphorylation.*

We begin our study of oxidative phosphorylation by examining the oxidation–reduction reactions that allow the flow of electrons from NADH and FADH$_2$ to oxygen. The electron flow takes place in four large protein complexes that are embedded in the inner mitochondrial membrane, together called the *respiratory chain* or the *electron-transport chain.*

$$\mathrm{NADH} + \tfrac{1}{2}\,\mathrm{O_2} + \mathrm{H^+} \longrightarrow \mathrm{H_2O} + \mathrm{NAD^+}$$
$$\Delta G^{\circ\prime} = -220.1\ \mathrm{kJ\ mol^{-1}}\ (-52.6\ \mathrm{kcal\ mol^{-1}})$$

The overall reaction is exergonic. Importantly, three of the complexes of the electron-transport chain use the energy released by the electron flow to pump protons out of the mitochondrial matrix. In essence, energy is transformed. The resulting unequal distribution of protons generates a pH gradient and a transmembrane electrical potential that creates a *proton-motive force*. ATP is synthesized when protons flow back to the mitochondrial matrix through an enzyme complex.

$$\mathrm{ADP} + \mathrm{P_i} + \mathrm{H^+} \rightarrow \mathrm{ATP} + \mathrm{H_2O}$$
$$\Delta G^{\circ\prime} = +30.5\ \mathrm{kJ\ mol^{-1}}\ (+7.3\ \mathrm{kcal\ mol^{-1}})$$

OUTLINE

18.1 Eukaryotic Oxidative Phosphorylation Takes Place in Mitochondria
18.2 Oxidative Phosphorylation Depends on Electron Transfer
18.3 The Respiratory Chain Consists of Four Complexes: Three Proton Pumps and a Physical Link to the Citric Acid Cycle
18.4 A Proton Gradient Powers the Synthesis of ATP
18.5 Many Shuttles Allow Movement Across Mitochondrial Membranes
18.6 The Regulation of Cellular Respiration Is Governed Primarily by the Need for ATP

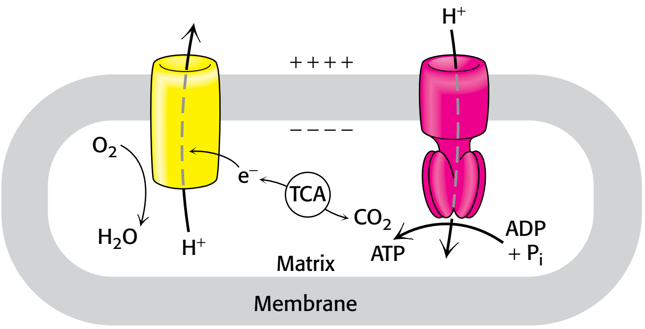


FIGURE 18.1 Overview of oxidative phosphorylation. Oxidation and ATP synthesis are coupled by transmembrane proton fluxes. Electrons flow from NADH and $FADH_2$ through four protein complexes to reduce oxygen to water (yellow tube). Three of the complexes pump protons from the mitochondrial matrix to the intermembrane space. The protons return to the matrix by flowing through another protein complex, ATP synthase (red structure), powering the synthesis of ATP.

Respiration

An ATP-generating process in which an inorganic compound (such as molecular oxygen) serves as the ultimate electron acceptor. The electron donor can be either an organic compound or an inorganic one.

Thus, *the oxidation of fuels and the phosphorylation of ADP are coupled by a proton gradient across the inner mitochondrial membrane* (Figure 18.1).

Collectively, the generation of high-transfer-potential electrons by the citric acid cycle, their flow through the respiratory chain, and the accompanying synthesis of ATP is called *respiration* or *cellular respiration*.

18.1 Eukaryotic Oxidative Phosphorylation Takes Place in Mitochondria

Recall that a biochemical role of the citric acid cycle, which takes place in the mitochondrial matrix, is the generation of high-energy electrons. It is fitting, therefore, that oxidative phosphorylation, which will convert the energy of these electrons into ATP, also takes place in mitochondria. Mitochondria are oval-shaped organelles, typically about 2 μm in length and 0.5 μm in diameter, about the size of a bacterium.

Mitochondria are bounded by a double membrane

Electron microscopic studies revealed that mitochondria have two membrane systems: an *outer membrane* and an extensive, highly folded *inner membrane.* The inner membrane is folded into a series of internal ridges called *cristae.* Hence, there are two compartments in mitochondria: (1) the *intermembrane space* between the outer and the inner membranes and (2) the *matrix,* which is bounded by the inner membrane (Figure 18.2). The mitochondrial matrix is the site of most of the reactions of the citric acid cycle and fatty acid oxidation. In contrast, oxidative phosphorylation takes place in the inner mitochondrial membrane. The increase in surface area of the inner mitochondrial membrane provided by the cristae creates more sites for oxidative phosphorylation than would be the case with a simple, unfolded membrane. Humans contain an estimated 14,000 m^2 of inner mitochondrial membrane, which is the approximate equivalent of three football fields in the United States.

The outer membrane is quite permeable to most small molecules and ions because it contains *mitochondrial porin,* a 30- to 35-kDa pore-forming protein also known as VDAC, for *v*oltage-*d*ependent *a*nion *c*hannel. VDAC, the most prevalent protein in the outer mitochondrial membrane, plays a role in the regulated flux of metabolites—usually anionic species such as phosphate, chloride, organic anions, and the adenine nucleotides—across the outer membrane. In contrast, the inner membrane is impermeable to nearly all ions and polar molecules. A large family of transporters shuttles metabolites such as ATP, pyruvate, and citrate across the inner mitochondrial membrane. The two faces of this membrane will be referred to as the *matrix side* and the *cytoplasmic side*

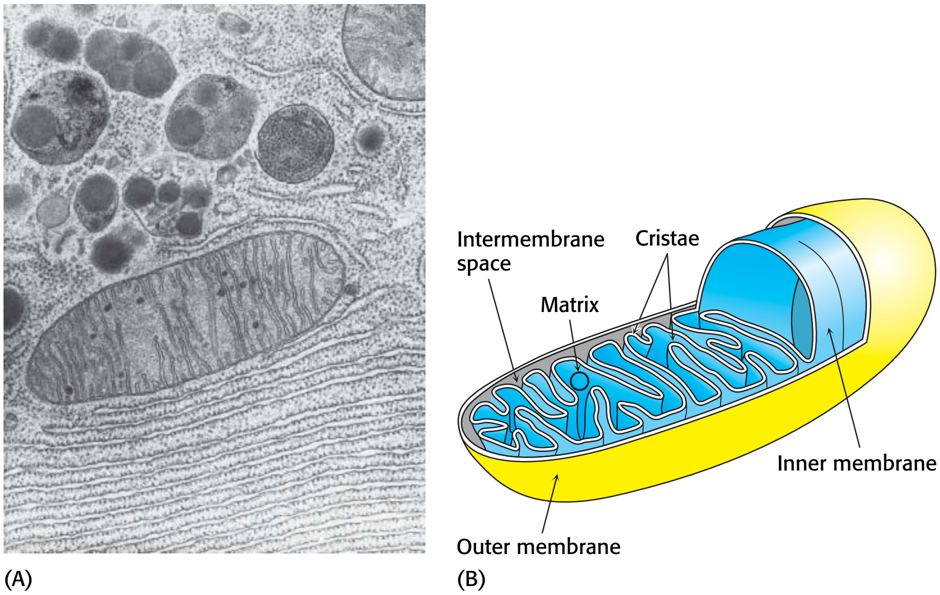


FIGURE 18.2 Electron micrograph (A) and diagram (B) of a mitochondrion. [(A) Keith R. Porter/Science Source. (B) Information from Wolfe, *Biology of the Cell,* 2e, © 1981 Brooks/Cole, a part of Cengage Learning, Inc. Reproduced by permission www. cengage.com/ permission 3.]

(the latter because it is freely accessible to most small molecules in the cytoplasm). They are also called the *N* and *P* sides, respectively, because the membrane potential is negative on the matrix side and positive on the cytoplasmic side.

In prokaryotes, the electron-driven proton pumps and ATP-synthesizing complexes are located in the cytoplasmic membrane, the inner of two membranes. The outer membrane of bacteria, like that of mitochondria, is permeable to most small metabolites because of the presence of porins.

Mitochondria are the result of an endosymbiotic event

Mitochondria are semiautonomous organelles that live in an endosymbiotic relation with the host cell. These organelles contain their own DNA, which encodes a variety of different proteins and RNAs. Mitochondrial DNA is usually portrayed as being circular, but recent research suggests that the mitochondrial DNA of many organisms may be linear. The genomes of mitochondria range broadly in size across species. The mitochondrial genome of the protist *Plasmodium falciparum* consists of fewer than 6000 base pairs (bp), whereas those of some land plants comprise more than 200,000 bp (Figure 18.3). Human mitochondrial DNA comprises 16,569 bp and encodes 13 respiratory-chain proteins as well as the small and large ribosomal RNAs and enough tRNAs to translate all codons. However, mitochondria also contain many proteins encoded by nuclear DNA. Cells that contain mitochondria depend on these organelles for oxidative phosphorylation, and the mitochondria in turn depend on the cell for their very existence. How did this intimate symbiotic relation come to exist?

An *endosymbiotic event* is thought to have occurred whereby a free-living organism capable of oxidative phosphorylation was engulfed by another cell. The double-membrane, circular DNA (with exceptions) and the mitochondrial-specific transcription and translation machinery all point to this conclusion. Thanks to the rapid accumulation of sequence data for mitochondrial and bacterial genomes, speculation on the origin of the "original" mitochondrion with some authority is now possible. The most mitochondrial-like bacterial genome is that of *Rickettsia prowazekii,* the cause of louse-borne typhus. The genome for this organism is more than 1 million base pairs in size and contains 834 protein-encoding genes.

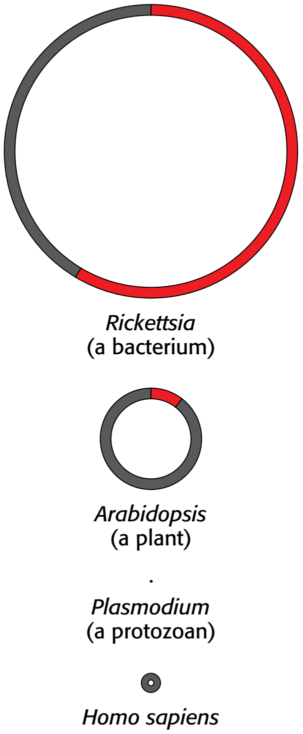


FIGURE 18.3 Sizes of mitochondrial genomes. The sizes of three mitochondrial genomes compared with the genome of *Rickettsia,* a relative of the presumed ancestor of all mitochondria. For genomes of more than 60 kbp, the DNA coding region for genes with known function is shown in red.

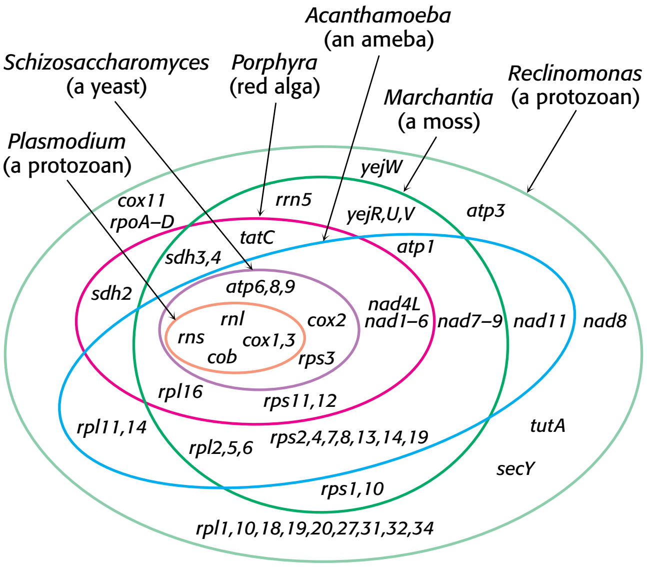

Acanthamoeba (an ameba)
Schizosaccharomyces (a yeast)
Porphyra (red alga)
Reclinomonas (a protozoan)
Marchantia (a moss)
Plasmodium (a protozoan)
yejW
cox11
rpoA–D
rrn5
yejR,U,V
atp3
tatC
atp1
sdh3,4
atp6,8,9
sdh2
nad4L
nad1–6
nad7–9
nad11
nad8
rnl
cox2
rns
cox1,3
cob
rps3
rpl16
rps11,12
rpl11,14
tutA
rpl2,5,6
rps2,4,7,8,13,14,19
secY
rps1,10
rpl1,10,18,19,20,27,31,32,34

FIGURE 18.4 Overlapping gene complements of mitochondria. The genes present within each oval are those present within the organism represented by the oval. Only rRNA- and protein-coding genes are shown. The genome of *Reclinomonas* contains all the protein-coding genes found in all the sequenced mitochondrial genomes. [Data from M. W. Gray, G. Burger, and B. F. Lang, *Science* 283:1476–1481, 1999.]

Sequence data suggest that all extant mitochondria are derived from an ancestor of *R. prowazekii* as the result of a single endosymbiotic event.

The evidence that modern mitochondria result from a single event comes from examination of the most bacteria-like mitochondrial genome, that of the protozoan *Reclinomonas americana*. Its genome contains 97 genes, of which 62 specify proteins. The genes encoding these proteins include all of the protein-coding genes found in all of the sequenced mitochondrial genomes (Figure 18.4). Yet, this genome encodes less than 2% of the protein-coding genes in the bacterium *E. coli*. In other words, a small fraction of bacterial genes—2%—is found in all examined mitochondria. How is it possible that all mitochondria have the same 2% of the bacterial genome? It seems unlikely that mitochondrial genomes resulting from several endosymbiotic events could have been independently reduced to the same set of genes found in *R. americana*. Thus, the simplest explanation is that the endosymbiotic event took place just once and all existing mitochondria are descendants of that ancestor.

Note that transient engulfment of prokaryotic cells by larger cells is not uncommon in the microbial world. In regard to mitochondria, such a transient relationship became permanent as the bacterial cell lost DNA, making it incapable of independent living, and the host cell became dependent on the ATP generated by its tenant.

18.2 Oxidative Phosphorylation Depends on Electron Transfer

In Chapter 17, the generation of NADH and $FADH_2$ by the oxidation of acetyl CoA was identified as an important function of the citric acid cycle. In oxidative phosphorylation, electrons from NADH and $FADH_2$ are used to reduce molecular oxygen to water. The highly exergonic reduction is accomplished by a number of electron-transfer reactions, which take place in a set of membrane proteins known as the *electron-transport chain*.

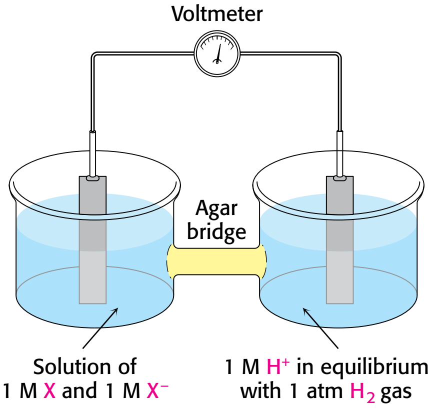

Voltmeter
Agar bridge
Solution of 1 M X and 1 M X^-
1 M H^+ in equilibrium with 1 atm H_2 gas

FIGURE 18.5 Measurement of redox potential. Apparatus for the measurement of the standard oxidation–reduction potential of a redox couple. Electrons flow through the wire connecting the cells, whereas ions flow through the agar bridge.

The electron-transfer potential of an electron is measured as redox potential

In oxidative phosphorylation, the *electron-transfer potential* of NADH or $FADH_2$ is converted into the *phosphoryl-transfer potential* of ATP. To better understand this conversion, we need quantitative expressions for these forms of free energy. The measure of phosphoryl-transfer potential is already familiar to us: it is given by $\Delta G^{\circ\prime}$ for the hydrolysis of the activated phosphoryl compound. The corresponding expression for the electron-transfer potential is E_0', the *reduction potential* (also called the *redox potential* or *oxidation–reduction potential*).

Consider a substance that can exist in an oxidized form X and a reduced form X^-. Such a pair is called a *redox couple* and is designated $X : X^-$. The reduction potential of this couple can be determined by measuring the electromotive force generated by an apparatus called a *sample half-cell* connected to a *standard reference half-cell* (Figure 18.5). The sample half-cell consists of an electrode immersed in a solution of 1 M oxidant (X) and 1 M reductant (X^-). The standard reference half-cell consists of an electrode immersed in a 1 M H^+ solution that is in equilibrium with H_2 gas at 1 atmosphere (1 atm) of pressure. The electrodes are connected to a voltmeter, and an agar bridge allows ions to move from one half-cell to the other, establishing electrical continuity between the half-cells. Electrons then flow from one half-cell to the other through the wire connecting the two half-cells to the voltmeter. If the reaction proceeds in the direction

$$X^- + H^+ \longrightarrow X + \tfrac{1}{2}H_2$$

the reactions in the half-cells (referred to as *half-reactions* or *couples*) must be

$$X^- \longrightarrow X + e^- \qquad H^+ + e^- \longrightarrow \tfrac{1}{2}H_2$$

Thus, electrons flow from the sample half-cell to the standard reference half-cell, and the sample-cell electrode is taken to be negative with respect to the standard-cell electrode. *The reduction potential of the $X : X^-$ couple is the observed voltage at the start of the experiment* (when X, X^-, and H^+ are 1 M with 1 atm of H_2). *The reduction potential of the $H^+ : H_2$* couple is defined to be 0 volts. In oxidation–reduction reactions, the donor of electrons, in this case X^-, is called the *reductant* or *reducing agent*, whereas the acceptor of electrons, H^+ here, is called the *oxidant* or *oxidizing agent*.

The meaning of the reduction potential is now evident. A negative reduction potential means that the oxidized form of a substance has lower affinity for electrons than does H_2, as in the preceding example. A positive reduction potential means that the oxidized form of a substance has higher affinity for electrons than does H_2. These comparisons refer to standard conditions—namely, 1 M oxidant, 1 M reductant, 1 M H^+, and 1 atm H_2. Thus, *a strong reducing agent (such as NADH) is poised to donate electrons and has a negative reduction potential, whereas a strong oxidizing agent (such as O_2) is ready to accept electrons and has a positive reduction potential.*

The reduction potentials of many biologically important redox couples are known (Table 18.1). Table 18.1 is like those presented in chemistry

TABLE 18.1 Standard reduction potentials of some reactions

Oxidant	Reductant	n	E'_0 (V)
Succinate + CO_2	α-Ketoglutarate	2	−0.67
Acetate	Acetaldehyde	2	−0.60
Ferredoxin (oxidized)	Ferredoxin (reduced)	1	−0.43
2 H^+	H_2	2	−0.42
NAD^+	NADH + H^+	2	−0.32
$NADP^+$	NADPH + H^+	2	−0.32
Lipoate (oxidized)	Lipoate (reduced)	2	−0.29
Glutathione (oxidized)	Glutathione (reduced)	2	−0.23
FAD	$FADH_2$	2	−0.22
Acetaldehyde	Ethanol	2	−0.20
Pyruvate	Lactate	2	−0.19
2 H^+	H_2	2	−0.00[1]
Fumarate	Succinate	2	+0.03
Cytochrome *b* (+3)	Cytochrome *b* (+2)	1	+0.07
Dehydroascorbate	Ascorbate	2	+0.08
Ubiquinone (oxidized)	Ubiquinone (reduced)	2	+0.10
Cytochrome *c* (+3)	Cytochrome *c* (+2)	1	+0.22
Fe (+3)	Fe (+2)	1	+0.77
$\tfrac{1}{2}O_2$ + 2 H^+	H_2O	2	+0.82

Note: E'_0 is the standard oxidation–reduction potential (pH 7, 25°C) and n is the number of electrons transferred. E'_0 refers to the partial reaction written as Oxidant + $e^- \longrightarrow$ reductant

[1] Standard oxidation − reduction potential at pH = 0.

textbooks, except that a hydrogen ion concentration of 10^{-7} M (pH 7) instead of 1 M (pH 0) is the standard state adopted by biochemists. This difference is denoted by the prime in E_0'. Recall that the prime in $\Delta G^{\circ\prime}$ denotes a standard free-energy change at pH 7.

The standard free-energy change $\Delta G^{\circ\prime}$ is related to the change in reduction potential $\Delta E_0'$ by

$$\Delta G^{\circ\prime} = -nF\Delta E_0'$$

in which n is the number of electrons transferred, F is a proportionality constant called the *Faraday constant* [96.48 kJ mol^{-1} V^{-1} (23.06 kcal mol^{-1} V^{-1})], $\Delta E_0'$ is in volts, and $\Delta G^{\circ\prime}$ is in kilojoules or kilocalories per mole.

The free-energy change of an oxidation–reduction reaction can be readily calculated from the reduction potentials of the reactants. For example, consider the reduction of pyruvate by NADH, catalyzed by lactate dehydrogenase. Recall that this reaction maintains redox balance in lactic acid fermentation (Figure 16.11).

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{lactate} + \text{NAD}^+ \qquad \text{(A)}$$

The reduction potential of the NAD$^+$: NADH couple, or half-reaction, is -0.32 V, whereas that of the pyruvate:lactate couple is -0.19 V. By convention, reduction potentials (as in Table 18.1) refer to partial reactions written as reductions: oxidant + e^- → reductant. Hence,

$$\begin{array}{llr}\text{Pyruvate} + 2\,\text{H}^+ + 2\,e^- \rightarrow \text{lactate} & E_0' = -0.19\text{ V} & \text{(B)}\\ \text{NAD}^+ + \text{H}^+ + 2\,e^- \rightarrow \text{NADH} & E_0' = -0.32\text{ V} & \text{(C)}\end{array}$$

To obtain reaction A from reactions B and C, we need to reverse the direction of reaction C so that NADH appears on the left side of the arrow. In doing so, the sign of E_0' must be changed.

$$\begin{array}{llr}\text{Pyruvate} + 2\,\text{H}^+ + 2\,e^- \rightarrow \text{lactate} & E_0' = -0.19\text{ V} & \text{(B)}\\ \text{NADH} \rightarrow \text{NAD}^+ + \text{H}^+ + 2\,e^- & E_0' = +0.32\text{ V} & \text{(D)}\end{array}$$

For reaction B, the free energy can be calculated with $n = 2$.

$$\begin{aligned}\Delta G^{\circ\prime} &= -2 \times 96.48\text{ kJ mol}^{-1}\text{ V}^{-1} \times -0.19\text{ V}\\ &= +36.7\text{ kJ mol}^{-1}(+8.8\text{ kcal mol}^{-1})\end{aligned}$$

Likewise, for reaction D,

$$\begin{aligned}\Delta G^{\circ\prime} &= -2 \times 96.48\text{ kJ mol}^{-1}\text{ V}^{-1} \times +0.32\text{ V}\\ &= -61.8\text{ kJ mol}^{-1}(-14.8\text{ kcal mol}^{-1})\end{aligned}$$

Thus, the free energy for reaction A is given by

$$\begin{aligned}\Delta G^{\circ\prime} &= \Delta G^{\circ\prime}(\text{for reaction B}) + \Delta G^{\circ\prime}(\text{for reaction D})\\ &= +36.7\text{ kJ mol}^{-1} - 61.8\text{ kJ mol}^{-1}\\ &= -25.1\text{ kJ mol}^{-1}(-6.0\text{ kcal mol}^{-1})\end{aligned}$$

A 1.14-volt potential difference between NADH and molecular oxygen drives electron transport through the chain and favors the formation of a proton gradient

The driving force of oxidative phosphorylation is the electron-transfer potential of NADH or FADH_2 relative to that of O_2. How much energy is released by the reduction of O_2 with NADH? Let us calculate $\Delta G^{\circ\prime}$ for this reaction. The pertinent half-reactions are

$$\tfrac{1}{2}\,O_2 + 2\,H^+ + 2\,e^- \longrightarrow H_2O \qquad E_0' = +0.82\,V \qquad (A)$$

$$NAD^+ + H^+ + 2\,e^- \longrightarrow NADH \qquad E_0' = -0.32\,V \qquad (B)$$

The combination of the two half-reactions, as it proceeds in the electron-transport chain, yields

$$\tfrac{1}{2}\,O_2 + NADH + H^+ \longrightarrow H_2O + NAD^+ \qquad (C)$$

The standard free energy for this reaction is then given by

$$\begin{aligned}\Delta G^{\circ\prime} &= (-2 \times 96.48\,\text{kJ mol}^{-1}\,\text{V}^{-1} \times +0.82\,\text{V}) \\ &\quad + (-2 \times 96.48\,\text{kJ mol}^{-1}\,\text{V}^{-1} \times 0.32\,\text{V}) \\ &= -158.2\,\text{kJ mol}^{-1} + (-61.9)\,\text{kJ mol}^{-1} \\ &= -220.1\,\text{kJ mol}^{-1}\,(-52.6\,\text{kcal mol}^{-1})\end{aligned}$$

This release of free energy is substantial. Recall that $\Delta G^{\circ\prime}$ for the synthesis of ATP is 30.5 kJ mol^{-1} (7.3 kcal mol^{-1}). The released energy is initially used to generate a proton gradient that is then used for the synthesis of ATP and the transport of metabolites across the mitochondrial membrane.

How can the energy associated with a proton gradient be quantified? Recall that the free-energy change for a species moving from one side of a membrane where it is at concentration c_1 to the other side where it is at a concentration c_2 is given by

$$\Delta G = RT\ln(c_2/c_1) + ZF\Delta V$$

in which Z is the electrical charge of the transported species and ΔV is the potential in volts across the membrane (Section 13.1). Under typical conditions for the inner mitochondrial membrane, the pH outside is 1.4 units lower than inside [corresponding to ln (c_2/c_1) of 3.2] and the membrane potential is 0.14 V, the outside being positive. Because $Z = +1$ for protons, the free-energy change is $(8.32 \times 10^{-3}\,\text{kJ mol}^{-1}\,\text{K}^{-1} \times 310\,\text{K} \times 3.2) + (+1 \times 96.48\,\text{kJ mol}^{-1}\,\text{V}^{-1} \times 0.14\,\text{V}) = 21.8\,\text{kJ mol}^{-1}$ (5.2 kcal mol^{-1}). Thus, each proton that is transported out of the matrix to the cytoplasmic side corresponds to 21.8 kJ mol^{-1} of free energy.

18.3 The Respiratory Chain Consists of Four Complexes: Three Proton Pumps and a Physical Link to the Citric Acid Cycle

Electrons are transferred from NADH to O_2 through a chain of three large protein complexes called *NADH-Q oxidoreductase, Q-cytochrome* c *oxidoreductase,* and *cytochrome* c *oxidase* (Figure 18.6 and Table 18.2). *Electron flow within these transmembrane complexes is highly exergonic and powers the transport of protons across the inner mitochondrial membrane.* A fourth large protein complex, called *succinate-Q reductase,* contains the succinate dehydrogenase that generates $FADH_2$ in the citric acid cycle. Electrons from this $FADH_2$ enter the electron-transport chain at Q-cytochrome *c* oxidoreductase. Succinate-Q reductase, in contrast with the other complexes, does not pump protons. NADH-Q oxidoreductase, succinate-Q reductase, Q-cytochrome *c* oxidoreductase, and cytochrome *c* oxidase are also called *Complex I, II, III,* and *IV,* respectively. Complexes I, III, and IV appear to be associated in a supramolecular complex termed the *respirasome*. As we have seen before, such supramolecular complexes facilitate the rapid transfer of substrate and prevent the release of reaction intermediates.

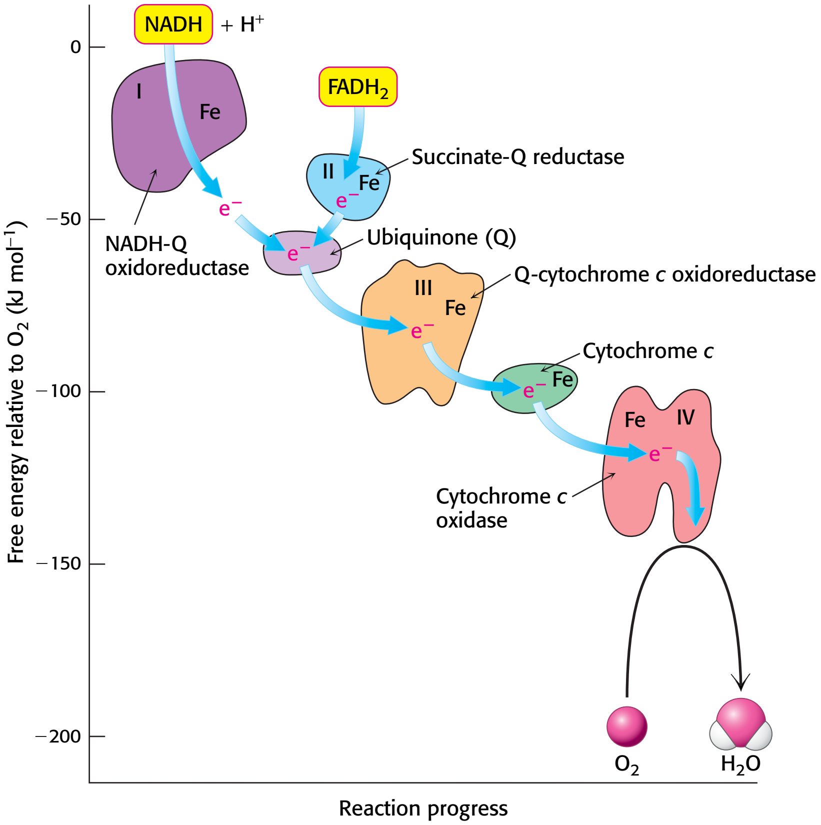

NADH
+ H+
I
Fe
NADH-Q oxidoreductase
FADH2
II
Fe
Succinate-Q reductase
Ubiquinone (Q)
III
Fe
Q-cytochrome c oxidoreductase
Cytochrome c
Fe
Fe
IV
Cytochrome c oxidase
e−
O2
H2O
Free energy relative to O2 (kJ mol−1)
0
−50
−100
−150
−200
Reaction progress

FIGURE 18.6 Components of the electron-transport chain. Electrons flow down an energy gradient from NADH to O_2. The flow is catalyzed by four protein complexes, and the energy released is used to generate a proton gradient. [Information from D. Sadava et al., *Life,* 8th ed. (Sinauer, 2008), p. 150.]

Two special electron carriers ferry the electrons from one complex to the next. The first is *coenzyme Q* (Q), also known as *ubiquinone* because it is a *ubi*quitous *quinone* in biological systems. Ubiquinone is a hydrophobic quinone that diffuses rapidly within the inner mitochondrial membrane. Electrons are carried from NADH-Q oxidoreductase to Q-cytochrome *c* oxidoreductase, the third complex of the chain, by the reduced form of Q. Electrons from the $FADH_2$ generated by the citric acid cycle are transferred first to ubiquinone and then to the Q-cytochrome *c* oxidoreductase complex.

Coenzyme Q is a quinone derivative with a long tail consisting of five-carbon isoprene units that account for its hydrophobic nature. The number of isoprene units in the tail depends on the species. The most common mammalian form contains 10 isoprene units (coenzyme Q_{10}). For simplicity, the subscript will be omitted from this abbreviation because all varieties

TABLE 18.2 Components of the mitochondrial electron-transport chain

				Oxidant or reductant		
Enzyme complex	Mass (kDa)	Subunits	Prosthetic group	Matrix side	Membrane core	Cytoplasmic side
NADH-Q oxidoreductase	>900	46	FMN Fe-S	NADH	Q	
Succinate-Q reductase	140	4	FAD Fe-S	Succinate	Q	
Q-cytochrome *c* oxidoreductase	250	11	Heme b_H Heme b_L Heme c_1 Fe-S		Q	Cytochrome *c*
Cytochrome *c* oxidase	160	13	Heme *a* Heme a_3 Cu_A and Cu_B			Cytochrome *c*

Information from: J. W. DePierre and L. Ernster, *Annu. Rev. Biochem.* 46:215, 1977; Y. Hatefi, *Annu Rev. Biochem.* 54:1015, 1985; and J. E. Walker, *Q. Rev. Biophys.* 25:253, 1992.

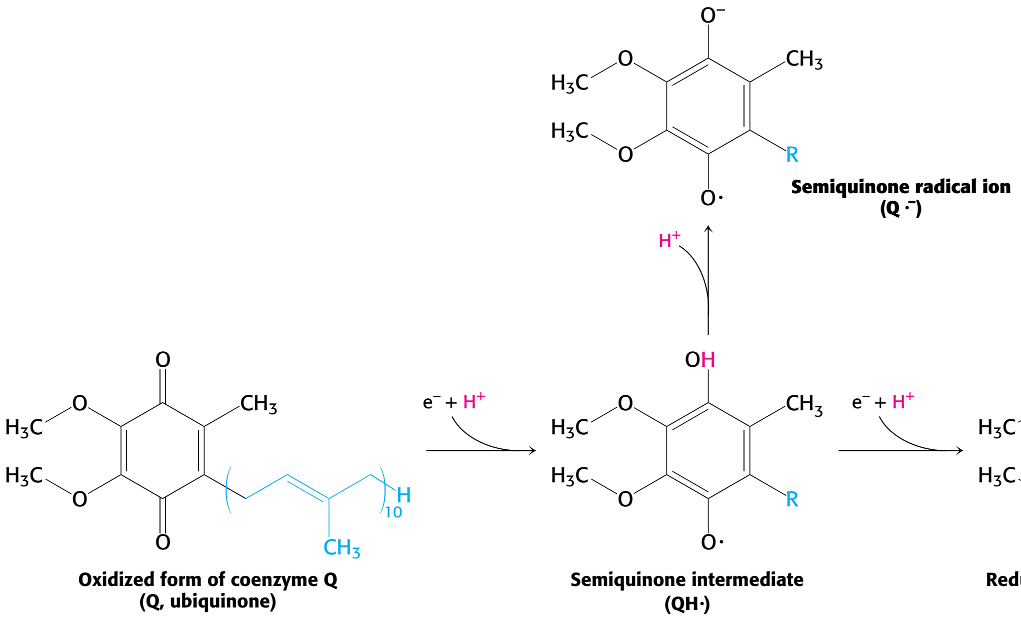


FIGURE 18.7 Oxidation states of quinones. The reduction of ubiquinone (Q) to ubiquinol (QH_2) proceeds through a semiquinone intermediate (QH·).

function in an identical manner. Quinones can exist in several oxidation states. In the fully oxidized state (Q), coenzyme Q has two keto groups (Figure 18.7). The addition of one electron and one proton results in the semiquinone form (QH·). The semiquinone can lose a proton to form a semiquinone radical anion ($Q\cdot^-$). The addition of a second electron and proton to the semiquinone generates ubiquinol (QH_2), the fully reduced form of coenzyme Q, which holds its protons more tightly. Thus, *for quinones, electron-transfer reactions are coupled to proton binding and release,* a property that is key to transmembrane proton transport. Because ubiquinone is soluble in the membrane, a pool of Q and QH_2—the *Q pool*—is thought to exist in the inner mitochondrial membrane, although recent research suggests that the Q pool is confined to the respirasome.

In contrast with Q, the second special electron carrier is a protein. Cytochrome *c,* a small soluble protein, shuttles electrons from Q-cytochrome *c* oxidoreductase to cytochrome *c* oxidase, the final component in the chain and the one that catalyzes the reduction of O_2.

Iron–sulfur clusters are common components of the electron transport chain

Iron–sulfur clusters in *iron–sulfur proteins* (also called *nonheme iron proteins*) play a critical role in a wide range of reduction reactions in biological systems. Several types of Fe-S clusters are known (Figure 18.8). In the simplest kind, a single iron ion is tetrahedrally coordinated to the sulfhydryl groups of four cysteine residues of the protein (Figure 18.8A). A second kind, denoted by 2Fe-2S, contains two iron ions, two inorganic sulfides, and usually four cysteine residues (Figure 18.8B). A third type, designated 4Fe-4S, contains four iron ions, four inorganic sulfides, and four cysteine residues (Figure 18.8C). NADH-Q oxidoreductase contains both 2Fe-2S and 4Fe-4S clusters. Iron ions in these Fe-S complexes cycle between Fe^{2+} (reduced) and Fe^{3+} (oxidized) states. Unlike quinones and flavins, iron–sulfur clusters generally undergo oxidation–reduction reactions without releasing or binding protons.

The importance of Fe-S clusters is illustrated by the loss of function of the protein frataxin. Frataxin is a small (14.2-kDa) mitochondrial protein that is crucial for the synthesis of Fe-S clusters. Mutations in frataxin causes Friedreich's ataxia, a disease of varying severity that affects

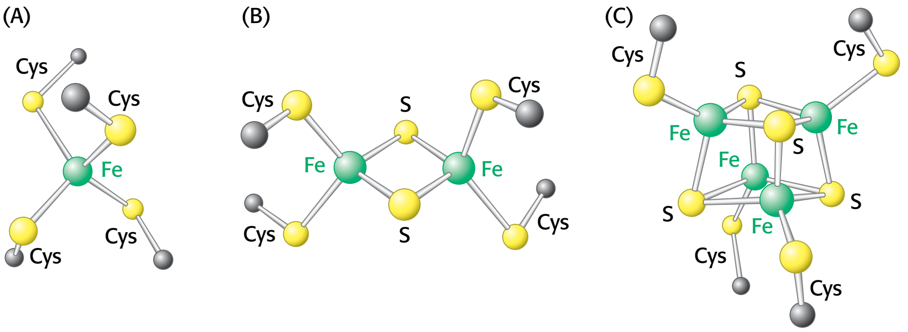


FIGURE 18.8 Iron–sulfur clusters. (A) A single iron ion bound by four cysteine residues. (B) 2Fe-2S cluster with iron ions bridged by sulfide ions. (C) 4Fe-4S cluster. Each of these clusters can undergo oxidation–reduction reactions.

the central and peripheral nervous as well as the heart and skeletal system. Severe cases lead to death in young adult life. The most common mutation is trinucleotide expansion (Section 28.5) in the gene for frataxin.

The high-potential electrons of NADH enter the respiratory chain at NADH-Q oxidoreductase

The electrons of NADH enter the chain at *NADH-Q oxidoreductase* (also called *Complex I* and *NADH dehydrogenase*), an enormous enzyme (>900 kDa) consisting of approximately 46 polypeptide chains. This proton pump, like that of the other two in the respiratory chain, is encoded by genes residing in both the mitochondria and the nucleus. NADH-Q oxidoreductase is L-shaped, with a horizontal arm lying in the membrane and a vertical arm that projects into the matrix.

The reaction catalyzed by this enzyme appears to be

$$\mathrm{NADH + Q + 5\,H^+_{matrix} \rightarrow NAD^+ + QH_2 + 4\,H^+_{cytoplasm}}$$

The initial step is the binding of NADH and the transfer of its two high-potential electrons to the *flavin mononucleotide* (FMN) prosthetic group, yielding the reduced form, $\mathrm{FMNH_2}$ (Figure 18.9). The electron acceptor of FMN, the isoalloxazine ring, is identical with that of FAD. Electrons are then transferred from $\mathrm{FMNH_2}$ to a series of iron–sulfur clusters, the second type of prosthetic group in NADH-Q oxidoreductase.

Recent structural studies have suggested how Complex I acts as a proton pump. What are the structural elements required for proton-pumping? The membrane-embedded part of the complex has four proton half-channels consisting, in part, of vertical helices. One set of half-channels is exposed to the matrix and the other to the inner membrane space (Figure 18.10). The

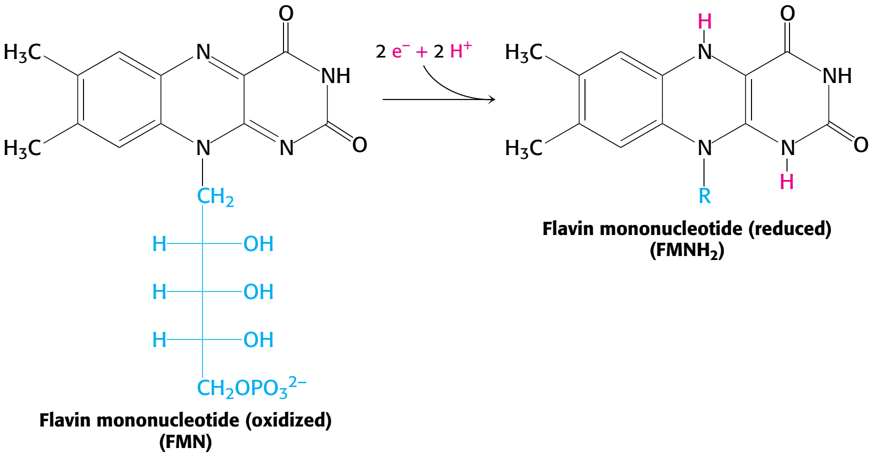

$2\,e^- + 2\,H^+$
Flavin mononucleotide (oxidized)
(FMN)
Flavin mononucleotide (reduced)
($FMNH_2$)

FIGURE 18.9 Oxidation states of flavins.

vertical helices are linked on the matrix side by a long horizontal helix (HL) that connects the matrix half-channels, while the cytoplasmic half-channels are joined by a series of β-hairpin-helix connecting elements (βH). An enclosed Q chamber, the site where Q accepts electrons from NADH, exists near the junction of the hydrophilic portion and the membrane-embedded portion. Finally, a hydrophilic funnel connects the Q chamber to a water-lined channel (into which the half channels open) that extends the entire length of the membrane-embedded portion.

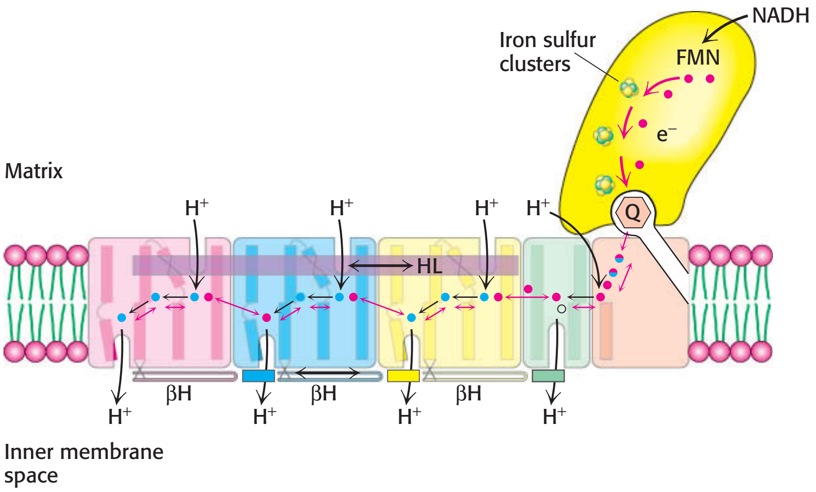
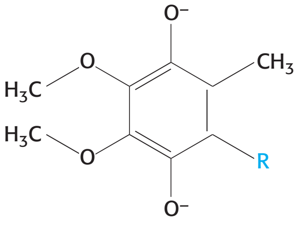


FIGURE 18.10 Coupled electron–proton transfer reactions through NADH-Q oxidoreductase. Electrons flow in Complex I from NADH through FMN and a series of iron–sulfur clusters to ubiquinone (Q), forming Q^{2-}. The charges on Q^{2-} are electrostatically transmitted to hydrophilic amino acid residues (shown as red (glutamate) and blue (lysine or histidine) balls) that power the movement of HL and βH components. This movement changes the conformation of the transmembrane helices and results in the transport of four protons out of the mitochondrial matrix. [Information from R. Baradaran et al., *Nature* 494:443–448.]

How do these structural elements cooperate to pump protons out of the matrix? When Q accepts two electrons from NADH, generating Q^{2-}, the negative charges on Q^{2-} interact electrostatically with negatively-charged amino acid residues in the membrane-embedded arm, causing conformational changes in the long horizontal helix and the βH elements. These changes in turn alter the structures of the connected vertical helices that change the pK_a of amino acids, allowing protons from the matrix to first bind to the amino acids, then dissociate into the water-lined channel and finally enter the intermembrane space. Thus, *the flow of two electrons from NADH to coenzyme Q through NADH-Q oxidoreductase leads to the pumping of four hydrogen ions out of the matrix of the mitochondrion.* Q^{2-} subsequently takes up two protons from the matrix as it is reduced to QH_2. The removal of these protons from the matrix contributes to the formation of the proton motive force. The QH_2 subsequently leaves the enzyme for the Q pool, allowing another reaction cycle to occur.



Reduced ubiquinone
(Q^{2-})

It is important to note that the citric acid cycle is not the only source of mitochondrial NADH. As we will see in Chapter 22, fatty acid degradation, which also takes place in mitochondria, is another crucial source of NADH for the electron-transport chain. Moreover, electrons from cytoplasmically generated NADH can be transported into mitochondria for use by the electron-transport chain (Section 18.5).

Ubiquinol is the entry point for electrons from $FADH_2$ of flavoproteins

$FADH_2$ enters the electron-transport chain at the second protein complex of the chain. Recall that $FADH_2$ is formed in the citric acid cycle, in the oxidation of succinate to fumarate by succinate dehydrogenase (Section 17.2). Succinate dehydrogenase is part of the *succinate-Q reductase complex (Complex II),* an integral membrane protein of the inner mitochondrial membrane. $FADH_2$ does not leave the complex. Rather, its electrons are transferred to Fe-S centers and then finally to Q to form QH_2, which then is ready to transfer electrons further down the electron-transport chain. The succinate-Q reductase complex, in contrast with NADH-Q oxidoreductase, does not pump protons from one side of the membrane to the other. Consequently, less ATP is formed from the oxidation of $FADH_2$ than from NADH.

Electrons flow from ubiquinol to cytochrome *c* through Q-cytochrome *c* oxidoreductase

What is the fate of ubiquinol generated by Complexes I and II? The electrons from QH_2 are passed on to *cytochrome* c (Cyt *c*), a water-soluble protein, by

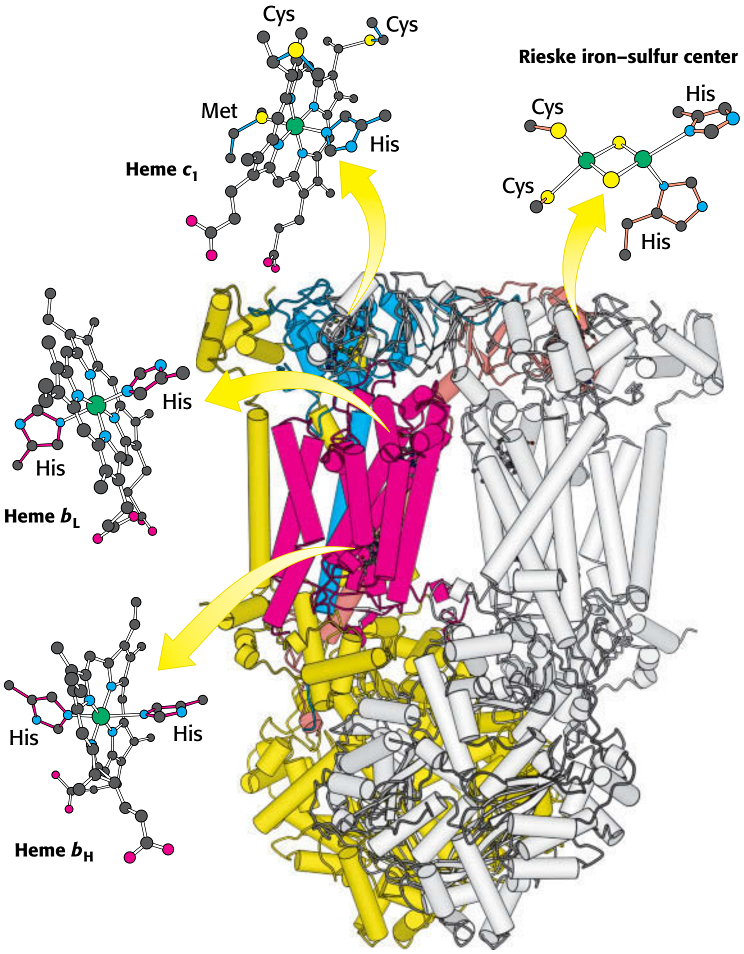


FIGURE 18.11 Structure of Q-cytochrome *c* oxidoreductase. This enzyme is a homodimer with each monomer consisting of 11 distinct polypeptide chains. Some of the more prominent components in one monomer are colored while the other monomer is white. Although each monomer contains the same components, some are identified in one monomer or the other for ease of viewing. *Notice* that the major prosthetic groups, three hemes and a 2Fe-2S cluster, are located either near the cytoplasmic edge of the complex bordering the intermembrane space (top) or in the region embedded in the membrane (α helices represented by tubes). They are well positioned to mediate the electron-transfer reactions between quinones in the membrane and cytochrome *c* in the intermembrane space. [Drawn from 1BCC.pdb.]

the second of the three proton pumps in the respiratory chain, *Q-cytochrome* c *oxidoreductase* (also known as *Complex III* and as *cytochrome* c *reductase*). The flow of a pair of electrons through this complex leads to the effective net transport of 2 H^+ to the cytoplasmic side, half the yield obtained with NADH-Q reductase because of a smaller thermodynamic driving force.

$$QH_2 + 2\,\text{Cyt}\,c_{ox} + 2\,H^+_{matrix} \rightarrow Q + 2\,\text{Cyt}\,c_{red} + 4\,H^+_{cytoplasm}$$

Q-cytochrome *c* oxidoreductase itself contains two types of cytochromes, named *b* and c_1 (Figure 18.11). *A cytochrome is an electron-transferring protein that contains a heme prosthetic group.* The heme prosthetic group in cytochromes *b*, c_1, and *c* is iron-protoporphyrin IX, the same heme present in myoglobin and hemoglobin (Section 7.1). The iron ion of a cytochrome, in contrast to hemoglobin and myoglobin, alternates between a reduced ferrous (+2) state and an oxidized ferric (+3) state during electron transport. The two cytochrome subunits of Q-cytochrome *c* oxidoreductase contain a total of three hemes: two hemes within cytochrome *b*, termed heme b_L (L for low affinity) and heme b_H (H for high affinity), and one heme within cytochrome c_1. These identical hemes have different electron affinities because they are in different polypeptide environments. For example, heme b_L, which is located in a cluster of helices near the cytoplasmic face of the membrane, has lower affinity for an electron than does heme b_H, which is near the matrix side.

In addition to the hemes, the enzyme contains an iron–sulfur protein with a 2Fe-2S center. This center, termed the *Rieske center,* is unusual in that one of the iron ions is coordinated by two histidine residues rather than two cysteine residues. This coordination stabilizes the center in its reduced form, raising its reduction potential so that it can readily accept electrons from QH_2.

The Q cycle funnels electrons from a two-electron carrier to a one-electron carrier and pumps protons

QH_2 passes two electrons to Q-cytochrome *c* oxidoreductase, but the acceptor of electrons in this complex, cytochrome *c*, can accept only one electron. How does the switch from the two-electron carrier ubiquinol to the one-electron carrier cytochrome *c* take place? The mechanism for the coupling of electron transfer from Q to cytochrome *c* to transmembrane proton transport is known as the *Q cycle* (Figure 18.12). Two QH_2 molecules bind to the complex consecutively, each giving up two electrons and two H^+. *These protons are released to the cytoplasmic side of the membrane.* The first QH_2 to exit the Q pool binds to the first Q binding site (Q_o), and its two electrons travel through the complex to different destinations. One electron flows, first, to the Rieske 2Fe-2S cluster; then, to cytochrome c_1; and, finally, to a molecule of oxidized cytochrome *c*, converting it into its reduced form. The reduced cytochrome *c* molecule is free to diffuse away from the enzyme to continue down the respiratory chain.

The second electron passes through two heme groups of cytochrome *b* to an oxidized ubiquinone in a second Q binding site (Q_i). The Q in the second binding site is reduced to a semiquinone radical anion ($Q\cdot^-$) by the electron from the first QH_2. The now fully oxidized Q leaves the first Q site, free to reenter the Q pool.

A second molecule of QH_2 binds to the Q_o site of Q-cytochrome *c* oxidoreductase and reacts in the same way as the first. One of the electrons is transferred to cytochrome *c* and the second electron is transferred to the partly reduced ubiquinone bound in the Q_i binding site. On the addition of the electron from the second QH_2 molecule, this quinone radical anion takes up two protons from the matrix side to form QH_2. *The removal of these two protons from the matrix contributes to the formation of the proton gradient*. In sum, four protons are released on the cytoplasmic side, and two protons are removed from the mitochondrial matrix.

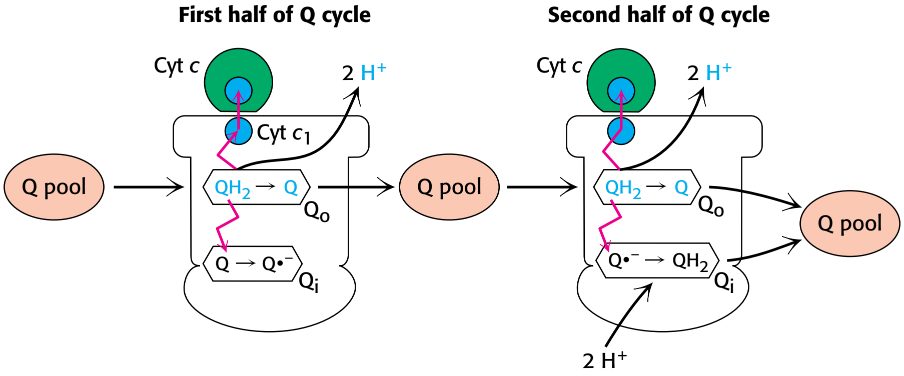


FIGURE 18.12 Q cycle. The Q cycle takes place in Complex III, which is represented in outline form. In the first half of the cycle, two electrons of a bound QH_2 are transferred, one to cytochrome *c* and the other to a bound Q in a second binding site to form the semiquinone radical anion $Q\cdot^-$. The newly formed Q dissociates and enters the Q pool. In the second half of the cycle, a second QH_2 also gives up its electrons to Complex III, one to a second molecule of cytochrome *c* and the other to reduce $Q\cdot^-$ to QH_2. This second electron transfer results in the uptake of two protons from the matrix. The path of electron transfer is shown in red.

$$2\,QH_2 + Q + 2\,\text{Cyt}\,c_{ox} + 2\,H^+_{matrix} \rightarrow 2\,Q + QH_2 + 2\,\text{Cyt}\,c_{red} + 4\,H^+_{cytoplasm}$$

In one Q cycle, two QH_2 molecules are oxidized to form two Q molecules, and then one Q molecule is reduced to QH_2. The problem of how to efficiently funnel electrons from a two-electron carrier (QH_2) to a one-electron carrier (cytochrome *c*) is solved by the Q cycle. The cytochrome *b* component of the reductase is in essence a recycling device that enables both electrons of QH_2 to be used effectively.

Cytochrome *c* oxidase catalyzes the reduction of molecular oxygen to water

The last of the three proton-pumping assemblies of the respiratory chain is *cytochrome* c *oxidase (Complex IV)*. Cytochrome *c* oxidase catalyzes the transfer of electrons from the reduced form of cytochrome *c* to molecular oxygen, the final acceptor.

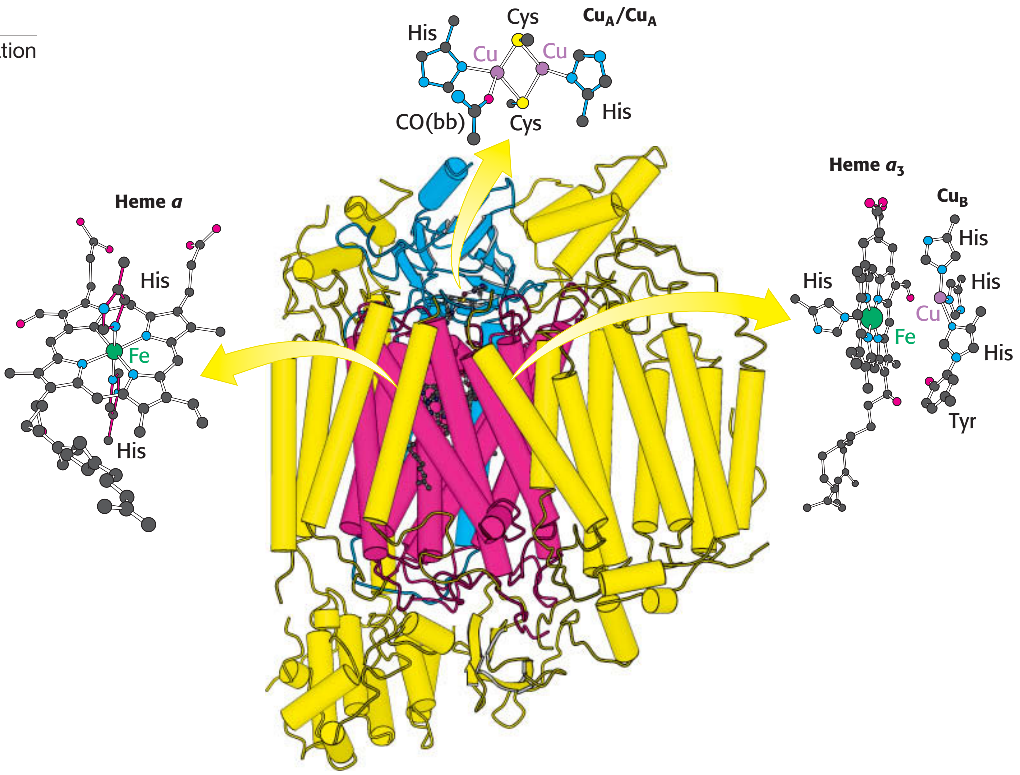
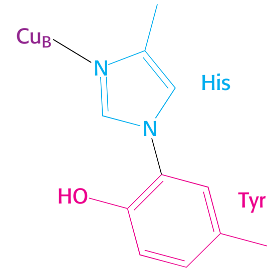


FIGURE 18.13 Structure of cytochrome *c* oxidase. This enzyme consists of 13 polypeptide chains. *Notice* that most of the complex, as well as two major prosthetic groups (heme *a* and heme a_3–Cu_B) are embedded in the membrane (α helices represented by vertical tubes). Heme a_3–Cu_B is the site of the reduction of oxygen to water. The Cu_A/Cu_A prosthetic group is positioned near the intermembrane space to better accept electrons from cytochrome *c*. CO(bb) is a carbonyl group of the peptide backbone. [Drawn from 2OCC.pdb.]

$$4\ \text{Cyt}\ c_{\text{red}} + 8\ \text{H}^+_{\text{matrix}} + \text{O}_2 \rightarrow 4\ \text{Cyt}\ c_{\text{ox}} + 2\ \text{H}_2\text{O} + 4\ \text{H}^+_{\text{cytoplasm}}$$

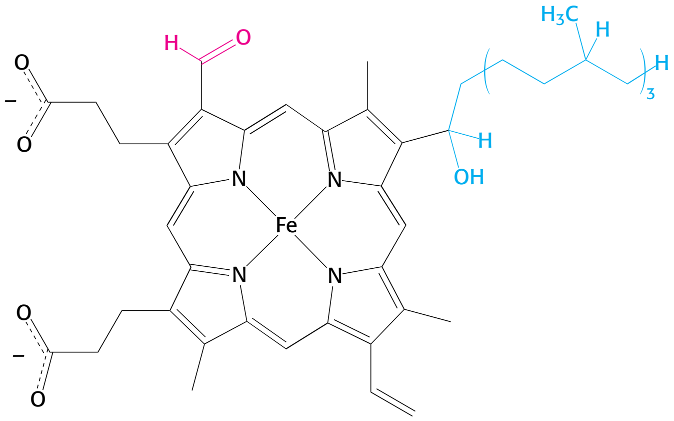
The requirement of oxygen for this reaction is what makes "aerobic" organisms aerobic. To obtain oxygen for this reaction is the reason that human beings must breathe. Four electrons are funneled to O_2 to completely reduce it to H_2O, and, concomitantly, protons are pumped from the matrix to the cytoplasmic side of the inner mitochondrial membrane. This reaction is quite thermodynamically favorable. From the reduction potentials in Table 18.1, the standard free-energy change for this reaction is calculated to be $\Delta G^{\circ\prime} = -231.8\ \text{kJ mol}^{-1}$ (-55.4 kcal mol^{-1}). As much of this free energy as possible must be captured in the form of a proton gradient for subsequent use in ATP synthesis.



Histidine covalently bound to tyrosine

Bovine cytochrome *c* oxidase is reasonably well understood at the structural level (Figure 18.13). It consists of 13 subunits, 3 of which are encoded by the mitochondrion's own genome. Cytochrome *c* oxidase contains two *heme A* groups and three *copper ions,* arranged as two copper centers, designated A and B. One center, Cu_A/Cu_A, contains two copper ions linked by two bridging cysteine residues. This center initially accepts electrons from reduced cytochrome *c*. The remaining copper ion, Cu_B, is bonded to three histidine residues, one of which is modified by covalent linkage to a tyrosine residue. The copper centers alternate between the reduced Cu^+ (cuprous) form and the oxidized Cu^{2+} (cupric) form as they accept and donate electrons.

There are two heme A molecules, called *heme* a and *heme* a_3, in cytochrome *c* oxidase. Heme A differs from the heme in cytochrome *c* and c_1 in three ways: (1) a formyl group replaces a methyl group, (2) a C_{17} hydrocarbon chain replaces one of the vinyl groups, and (3) the heme is not covalently attached to the protein.


H
O
H_3C
H
H
3
H
OH
O
–
O
N
N
Fe
N
N
O
–
O

Heme A

Heme *a* and heme a_3 have distinct redox potentials because they are located in different environments within cytochrome *c* oxidase. An electron flows from cytochrome *c* to Cu_A/Cu_A, to heme *a*, to heme a_3, to Cu_B, and finally to O_2. Heme a_3 and Cu_B are directly adjacent. Together, heme a_3 and Cu_B form the active center at which O_2 is reduced to H_2O.

Four molecules of cytochrome *c* bind consecutively to the enzyme and transfer an electron to reduce one molecule of O_2 to H_2O (Figure 18.14):

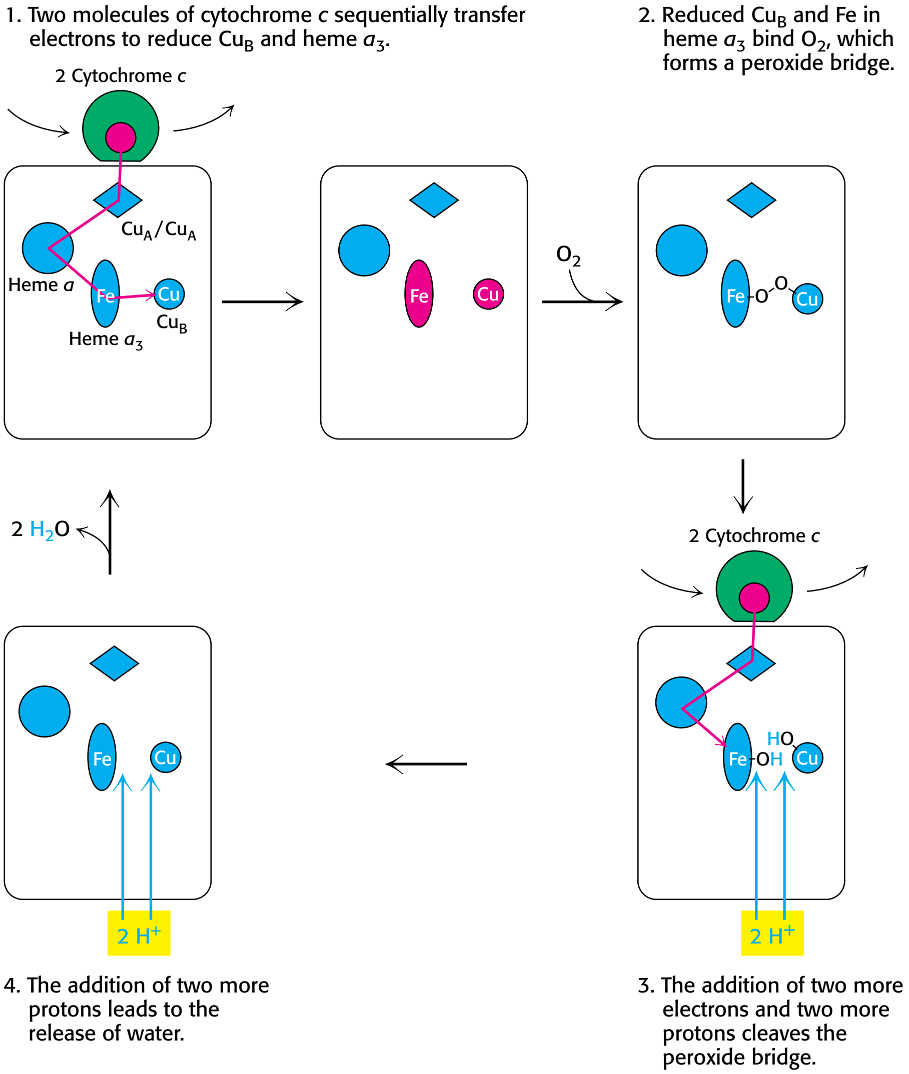

1. Two molecules of cytochrome c sequentially transfer electrons to reduce Cu_B and heme a_3.
2 Cytochrome c
Cu_A/Cu_A
Heme a
Fe
Cu
Cu_B
Heme a_3
Fe
Cu
O_2
2. Reduced Cu_B and Fe in heme a_3 bind O_2, which forms a peroxide bridge.
Fe
O
O
Cu
2 Cytochrome c
HO
Fe
OH
Cu
2 H^+
3. The addition of two more electrons and two more protons cleaves the peroxide bridge.
2 H_2O
Fe
Cu
2 H^+
4. The addition of two more protons leads to the release of water.

FIGURE 18.14 Cytochrome *c* oxidase mechanism. The cycle begins and ends with all prosthetic groups in their oxidized forms (shown in blue). Reduced forms are in red. Four cytochrome *c* molecules donate four electrons, which, in allowing the binding and cleavage of an O_2 molecule, also makes possible the import of four H^+ from the matrix to form two molecules of H_2O, which are released from the enzyme to regenerate the initial state.

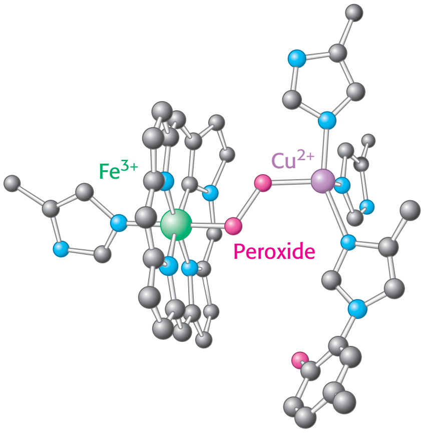


FIGURE 18.15 Peroxide bridge. The oxygen bound to heme a_3 is reduced to peroxide by the presence of Cu_B.

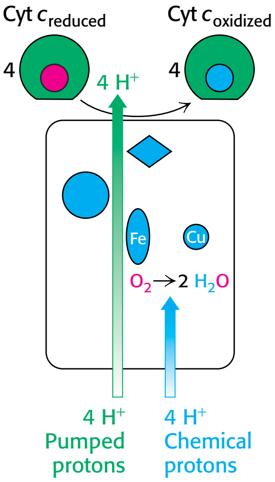


FIGURE 18.16 Proton transport by cytochrome *c* oxidase. Four protons are taken up from the matrix side to reduce one molecule of O_2 to two molecules of H_2O. These protons are called "chemical protons" because they participate in a clearly defined reaction with O_2. Four additional "pumped" protons are transported out of the matrix and released on the cytoplasmic side in the course of the reaction. The pumped protons double the efficiency of free-energy storage in the form of a proton gradient for this final step in the electron-transport chain.

1. Electrons from two molecules of reduced cytochrome *c* flow down an electron-transfer pathway within cytochrome *c* oxidase, one stopping at Cu_B and the other at heme a_3. With both centers in the reduced state, they together can now bind an oxygen molecule.

2. As molecular oxygen binds, it abstracts an electron from each of the nearby ions in the active center to form a peroxide (O_2^{2-}) bridge between them (Figure 18.15).

3. Two more molecules of cytochrome *c* bind and release electrons that travel to the active center. The addition of an electron as well as H^+ to each oxygen atom reduces the two ion–oxygen groups to Cu_B^{2+}—OH and Fe^{3+}—OH.

4. Reaction with two more H^+ ions allows the release of two molecules of H_2O and resets the enzyme to its initial, fully oxidized form.

$$4\ \text{Cyt}\ c_{\text{red}} + 4\ \text{H}^+_{\text{matrix}} + \text{O}_2 \rightarrow 4\ \text{Cyt}\ c_{\text{ox}} + 2\ \text{H}_2\text{O}$$

The four protons in this reaction come exclusively from the matrix. Thus, the consumption of these four protons contributes directly to the proton gradient. Recall that each proton contributes 21.8 kJ mol^{-1} (5.2 kcal mol^{-1}) to the free energy associated with the proton gradient; so these four protons contribute 87.2 kJ mol^{-1} (20.8 kcal mol^{-1}), an amount substantially less than the free energy available from the reduction of oxygen to water. What is the fate of this missing energy? Remarkably, *cytochrome* c *oxidase uses this energy to pump four additional protons from the matrix to the cytoplasmic side of the membrane in the course of each reaction cycle for a total of eight protons removed from the matrix* (Figure 18.16). The details of how these protons are transported through the protein are still under study. However, two effects contribute to the mechanism. First, charge neutrality tends to be maintained in the interior of proteins. Thus, the addition of an electron to a site inside a protein tends to favor the binding of H^+ to a nearby site. Second, conformational changes take place, particularly around the heme a_3–Cu_B center, in the course of the reaction cycle. Presumably, in one conformation, protons may enter the protein exclusively from the matrix side, whereas, in another, they may exit exclusively to the cytoplasmic side. Thus, the overall process catalyzed by cytochrome *c* oxidase is

$$4\ \text{Cyt}\ c_{\text{red}} + 8\ \text{H}^+_{\text{matrix}} + \text{O}_2 \rightarrow 4\ \text{Cyt}\ c_{\text{ox}} + 2\ \text{H}_2\text{O} + 4\ \text{H}^+_{\text{cytoplasm}}$$

Figure 18.17 summarizes the flow of electrons from NADH and $FADH_2$ through the respiratory chain. This series of exergonic reactions is coupled to the pumping of protons from the matrix. As we will see shortly, the energy inherent in the proton gradient will be used to synthesize ATP.

Toxic derivatives of molecular oxygen such as superoxide radicals are scavenged by protective enzymes

As discussed earlier, molecular oxygen is an ideal terminal electron acceptor, because its high affinity for electrons provides a large thermodynamic driving force. However, danger lurks in the reduction of O_2. The transfer of four electrons leads to safe products (two molecules of H_2O), but partial reduction generates hazardous compounds. In particular, *the transfer of a single electron to O_2 forms superoxide ion, whereas the transfer of two electrons yields peroxide.*

$$\text{O}_2 \xrightarrow{e^-} \underset{\text{Superoxide ion}}{\text{O}_2^{\cdot -}} \xrightarrow{e^-} \underset{\text{Peroxide}}{\text{O}_2^{2-}}$$

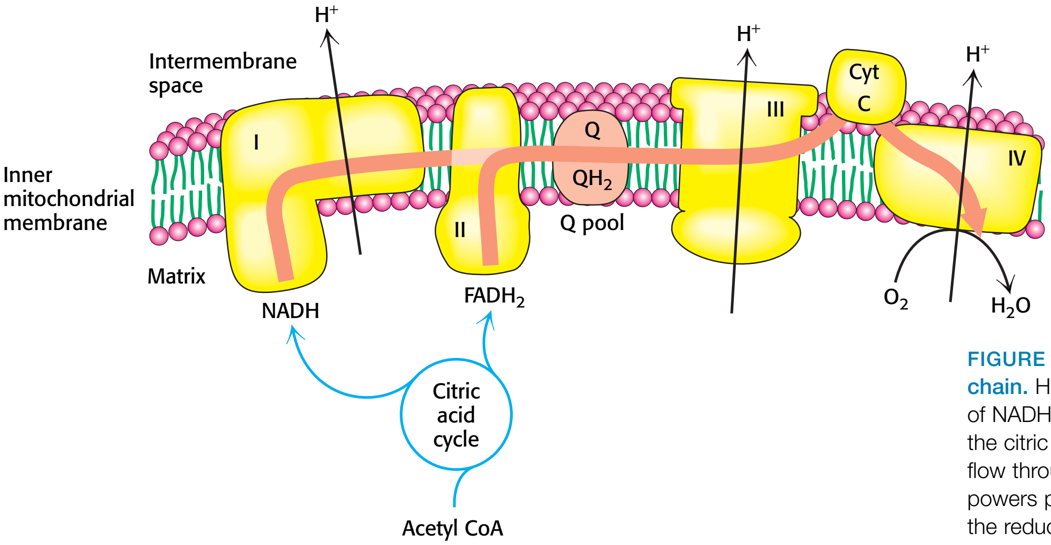


FIGURE 18.17 The electron-transport chain. High-energy electrons in the form of NADH and $FADH_2$ are generated by the citric acid cycle. These electrons flow through the respiratory chain, which powers proton pumping and results in the reduction of O_2.

Both compounds are potentially destructive. The strategy for the safe reduction of O_2 is clear: *the catalyst does not release partly reduced intermediates.* Cytochrome *c* oxidase meets this crucial criterion by holding O_2 tightly between Fe and Cu ions.

Although cytochrome *c* oxidase and other proteins that reduce O_2 are remarkably successful in not releasing intermediates, small amounts of superoxide anion and hydrogen peroxide are unavoidably formed. Superoxide, hydrogen peroxide, and species that can be generated from them such as the hydroxyl radical (OH·) are collectively referred to as *reactive oxygen species* or *ROS.* Oxidative damage caused by ROS has been implicated in the aging process as well as in a growing list of diseases (Table 18.3).

What are the cellular defense strategies against oxidative damage by ROS? Chief among them is the enzyme *superoxide dismutase.* This enzyme scavenges superoxide radicals by catalyzing the conversion of two of these radicals into hydrogen peroxide and molecular oxygen.

Dismutation
A reaction in which a single reactant is converted into two different products.

$$O_2^{\dot{-}} + 2H^+ \xrightleftharpoons{\text{Superoxide dismutase}} O_2 + H_2O_2$$

Eukaryotes contain two forms of this enzyme, a manganese-containing version located in mitochondria and a copper-and-zinc-dependent cytoplasmic form. These enzymes perform the dismutation reaction by a similar

TABLE 18.3 Pathological conditions that may entail free-radical injury

Atherogenesis
Emphysema; bronchitis
Parkinson disease
Duchenne muscular dystrophy
Cervical cancer
Alcoholic liver disease
Diabetes
Acute renal failure
Down syndrome
Retrolental fibroplasia (conversion of the retina into a fibrous mass in premature infants)
Cerebrovascular disorders
Ischemia; reperfusion injury

Information from Michael Lieberman and Allan D. Marks, *Basic Medical Biochemistry: A Clinical Approach,* 4th ed. (Lippincott, Williams & Wilkins, 2012), p. 437.

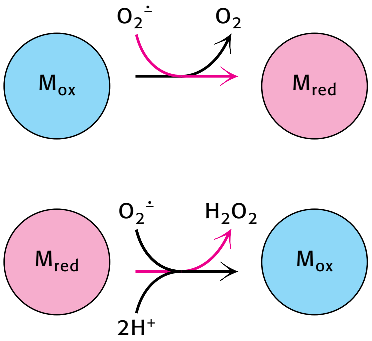


FIGURE 18.18 Superoxide dismutase mechanism. The oxidized form of superoxide dismutase (M_{ox}) reacts with one superoxide ion to form O_2 and generate the reduced form of the enzyme (M_{red}). The reduced form then reacts with a second superoxide and two protons to form hydrogen peroxide and regenerate the oxidized form of the enzyme.

mechanism (Figure 18.18). The oxidized form of the enzyme is reduced by superoxide to form oxygen. The reduced form of the enzyme, formed in this reaction, then reacts with a second superoxide ion to form peroxide, which takes up two protons along the reaction path to yield hydrogen peroxide.

The hydrogen peroxide formed by superoxide dismutase and by other processes is scavenged by *catalase,* a ubiquitous heme protein that catalyzes the dismutation of hydrogen peroxide into water and molecular oxygen.

$$2\,H_2O_2 \xrightleftharpoons{\text{Catalase}} O_2 + 2\,H_2O$$

Superoxide dismutase and catalase are remarkably efficient, performing their reactions at or near the diffusion-limited rate (Section 8.4). Glutathione peroxidase also plays a role in scavenging H_2O_2 (Section 20.5). Other cellular defenses against oxidative damage include the antioxidant vitamins, vitamins E and C. Because it is lipophilic, vitamin E is especially useful in protecting membranes from lipid peroxidation.

A long-term benefit of exercise may be to increase the amount of superoxide dismutase in the cell. The elevated aerobic metabolism during exercise causes more ROS to be generated. In response, the cell synthesizes more protective enzymes. The net effect is one of protection, because the increase in superoxide dismutase more effectively protects the cell during periods of rest (Problem 48).

Despite the fact that reactive oxygen species are known hazards, recent evidence suggests that, under certain circumstances, the controlled generation of these molecules may be an important component of signal-transduction pathways. For instance, growth factors have been shown to increase ROS levels as part of their signaling pathway, and ROS regulate channels and transcription factors. ROS have been implicated in the control of cell differentiation, the immune response, autophagy as well as other metabolic activities. The dual role of ROS is an excellent example of the wondrous complexity of biochemistry of living systems: even potentially harmful substances can be harnessed to play useful roles.

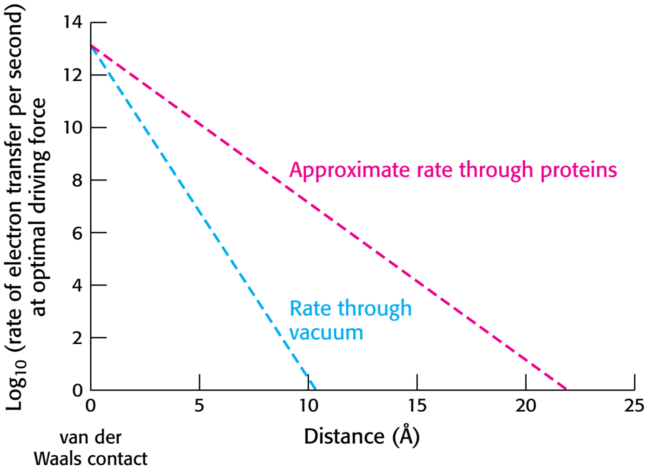


FIGURE 18.19 Distance dependence of electron-transfer rate. The rate of electron transfer decreases as the electron donor and the electron acceptor move apart. In a vacuum, the rate decreases by a factor of 10 for every increase of 0.8 Å. In proteins, the rate decreases more gradually, by a factor of 10 for every increase of 1.7 Å. This rate is only approximate because variations in the structure of the intervening protein medium can affect the rate.

Electrons can be transferred between groups that are not in contact

How are electrons transferred between electron-carrying groups of the respiratory chain? This question is intriguing because these groups are frequently buried in the interior of a protein in fixed positions and are therefore not directly in contact with one another. Electrons can move through space, even through a vacuum. However, the rate of electron transfer through space falls off rapidly as the electron donor and electron acceptor move apart from each other, decreasing by a factor of 10 for each increase in separation of 0.8 Å. The protein environment provides more-efficient pathways for electron conduction: typically, the rate of electron transfer decreases by a factor of 10 every 1.7 Å (Figure 18.19). For groups in contact, electron-transfer reactions can be quite fast, with rates of approximately $10^{13}\ s^{-1}$. Within proteins in the electron-transport chain, electron-carrying groups are typically separated by 15 Å beyond their van der Waals contact distance. For such separations, we expect electron-transfer rates of approximately $10^4\ s^{-1}$ (i.e., electron transfer in less than 1 ms), assuming that all other factors are optimal. Without the mediation of the protein, an electron transfer over this distance would take approximately 1 day.

The case is more complicated when electrons must be transferred between two distinct proteins, such as when cytochrome *c* accepts electrons from Complex III or passes them on to Complex IV. A series of hydrophobic interactions bring the heme groups of cytochrome *c* and c_1 to within 4.5 Å of each other, with the iron atoms separated by 17.4 Å. This distance could allow cytochrome *c* reduction at a rate of $8.3 \times 10^6\ \text{s}^{-1}$.

The conformation of cytochrome *c* has remained essentially constant for more than a billion years

Cytochrome *c* is present in all organisms having mitochondrial respiratory chains: plants, animals, and eukaryotic microorganisms. This electron carrier evolved more than 1.5 billion years ago, before the divergence of plants and animals. Its function has been conserved throughout this period, as evidenced by the fact that *the cytochrome* c *of any eukaryotic species reacts in vitro with the cytochrome* c *oxidase of any other species tested thus far.* For example, wheat-germ cytochrome *c* reacts with human cytochrome *c* oxidase. Additionally, some prokaryotic cytochromes, such as cytochrome c_2 from the photosynthetic bacterium *Rhodospirillum rubrum* and cytochrome c_{550} from the denitrifying bacterium *Paracoccus denitrificans*, closely resemble cytochrome *c* from tuna-heart mitochondria (Figure 18.20). This evidence attests to an efficient evolutionary solution to electron transfer bestowed by the structural and functional characteristics of cytochrome *c*.

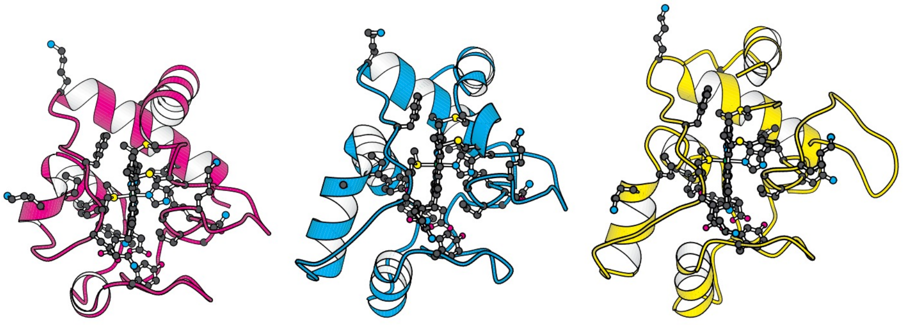

Tuna
Rhodospirillum rubrum
Paracoccus denitrificans

FIGURE 18.20 Conservation of the three-dimensional structure of cytochrome *c*. *Notice* the overall structural similarity of the different molecules from different sources. The side chains are shown for the 21 conserved amino acids as well as the centrally located planar heme. [Drawn from 3CYT.pdb, 3C2C.pdb, and 155C.pdb.]

The resemblance among cytochrome *c* molecules extends to the level of amino acid sequence. Because of the molecule's small size and ubiquity, the amino acid sequences of cytochrome *c* from more than 80 widely ranging eukaryotic species have been determined by direct protein sequencing. The striking finding is that *21 of 104 residues have been invariant for more than one and a half billion years of evolution.* A phylogenetic tree, constructed from the amino acid sequences of cytochrome *c*, reveals the evolutionary relationships between many animal species (Figure 18.21).

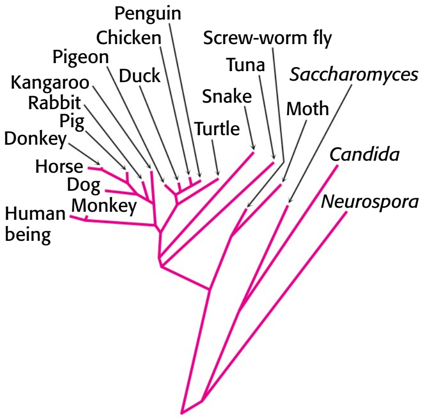


FIGURE 18.21 Evolutionary tree constructed from sequences of cytochrome *c*. Branch lengths are proportional to the number of amino acid changes that are believed to have occurred. [Information from Walter M. Fitch and Emanuel Margoliash.]

18.4 A Proton Gradient Powers the Synthesis of ATP

Thus far, we have considered the flow of electrons from NADH to O_2, an exergonic process.

$$\text{NADH} + \tfrac{1}{2}\,\text{O}_2 + \text{H}^+ \rightleftharpoons \text{H}_2\text{O} + \text{NAD}^+$$
$$\Delta G^{\circ\prime} = -220.1\ \text{kJ mol}^{-1}\ (-52.6\ \text{kcal mol}^{-1})$$

Next, we consider how this process is coupled to the synthesis of ATP, an endergonic process.

$$\mathrm{ADP} + \mathrm{P_i} + \mathrm{H^+} \rightleftharpoons \mathrm{ATP} + \mathrm{H_2O}$$
$$\Delta G^{\circ\prime} = +30.5\ \mathrm{kJ\ mol^{-1}}\,(+7.3\ \mathrm{kcal\ mol^{-1}})$$

A molecular assembly in the inner mitochondrial membrane carries out the synthesis of ATP. This enzyme complex was originally called the *mitochondrial ATPase* or *F_1F_0 ATPase* because it was discovered through its catalysis of the reverse reaction, the hydrolysis of ATP. *ATP synthase,* its preferred name, emphasizes its actual role in the mitochondrion. It is also called *Complex V*.

How is the oxidation of NADH coupled to the phosphorylation of ADP? Electron transfer was first suggested to lead to the formation of a covalent high-energy intermediate that serves as a compound having a high phosphoryl-transfer potential, analogous to the generation of ATP by the formation of 1,3-bisphosphoglycerate in glycolysis (Section 16.1). An alternative proposal was that electron transfer aids the formation of an activated protein conformation, which then drives ATP synthesis. The search for such intermediates for several decades proved fruitless.

In 1961, Peter Mitchell suggested a radically different mechanism, *the chemiosmotic hypothesis*. He proposed that electron transport and ATP synthesis are coupled by *a proton gradient across the inner mitochondrial membrane*. In his model, the transfer of electrons through the respiratory chain leads to the pumping of protons from the matrix to the cytoplasmic side of the inner mitochondrial membrane. The H^+ concentration becomes lower in the matrix, and an electric field with the matrix side negative is generated (Figure 18.22). Protons then flow back into the matrix to equalize the distribution. Mitchell's idea was that this flow of protons drives the synthesis of ATP by ATP synthase. The energy-rich unequal distribution of protons is called the *proton-motive force*. The proton-motive force is composed of two components: a chemical gradient and a charge gradient. The chemical gradient for protons can be represented as a pH gradient. The charge gradient is created by the positive charge on the unequally distributed protons forming the chemical gradient. Mitchell proposed that both components power the synthesis of ATP.

Some have argued that, along with the elucidation of the structure of DNA, the discovery that ATP synthesis is powered by a proton gradient is one of the two major advances in biology in the twentieth century. Mitchell's initial postulation of the chemiosmotic theory was not warmly received by all. Efraim Racker, one of the early investigators of ATP synthase, recalls that some thought of Mitchell as a court jester, whose work was of no consequence. Peter Mitchell was awarded the Nobel Prize in chemistry in 1978 for his contributions to understanding oxidative phosphorylation.

$$\text{Proton-motive force}(\Delta p) = \text{chemical gradient}(\Delta pH) + \text{charge gradient}(\Delta\psi)$$

Mitchell's highly innovative hypothesis that oxidation and phosphorylation are coupled by a proton gradient is now supported by a wealth of evidence. Indeed, electron transport does generate a proton gradient across the inner mitochondrial membrane. The pH outside is 1.4 units lower than inside, and the membrane potential is 0.14 V, the outside being positive. As calculated on page 529, this membrane potential corresponds to a free energy of 21.8 kJ (5.2 kcal) per mole of protons.

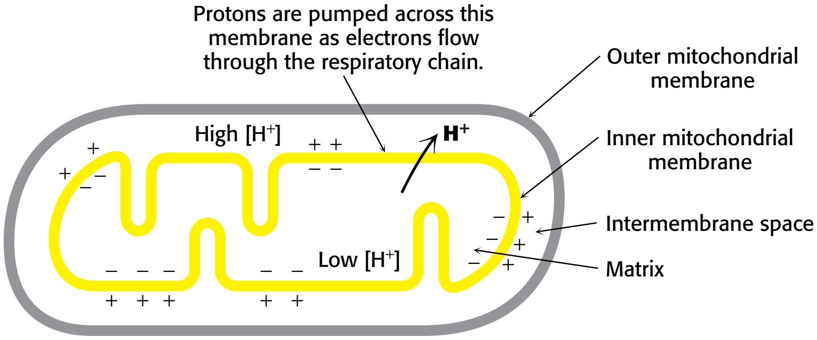

Protons are pumped across this membrane as electrons flow through the respiratory chain.
Outer mitochondrial membrane
Inner mitochondrial membrane
Intermembrane space
Matrix
High [H⁺]
Low [H⁺]
H⁺

FIGURE 18.22 Chemiosmotic hypothesis. Electron transfer through the respiratory chain leads to the pumping of protons from the matrix to the cytoplasmic side of the inner mitochondrial membrane. The pH gradient and membrane potential constitute a proton-motive force that is used to drive ATP synthesis.

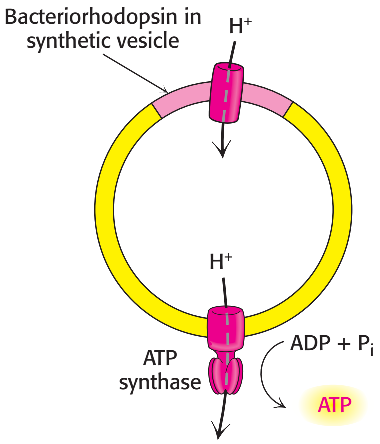


FIGURE 18.23 Testing the chemiosmotic hypothesis. ATP is synthesized when reconstituted membrane vesicles containing bacteriorhodopsin (a light-driven proton pump) and ATP synthase are illuminated.

An artificial system was created to elegantly demonstrate the basic principle of the chemiosmotic hypothesis. The role of the respiratory chain was played by bacteriorhodopsin, a membrane protein from halobacteria that pumps protons when illuminated. Synthetic vesicles containing bacteriorhodopsin and mitochondrial ATP synthase purified from beef heart were created (Figure 18.23). When the vesicles were exposed to light, ATP was formed. This key experiment clearly showed that *the respiratory chain and ATP synthase are biochemically separate systems, linked only by a proton-motive force.*

ATP synthase is composed of a proton-conducting unit and a catalytic unit

Two parts of the puzzle of how NADH oxidation is coupled to ATP synthesis are now evident: (1) electron transport generates a proton-motive force, and (2) ATP synthesis by ATP synthase can be powered by a proton-motive force. How is the proton-motive force converted into the high phosphoryl-transfer potential of ATP?

Biochemical, electron microscopic, and crystallographic studies of ATP synthase have revealed many details of its structure (Figure 18.24). It is a large, complex enzyme resembling a ball on a stick. Much of the "stick" part, called the F_0 subunit, is embedded in the inner mitochondrial membrane. The 85-Å-diameter ball, called the F_1 subunit, protrudes into the mitochondrial matrix. The F_1 subunit contains the catalytic activity of the synthase. In fact, isolated F_1 subunits display ATPase activity.

The F_1 subunit consists of five types of polypeptide chains (α_3, β_3, γ, δ, and ε) with the indicated stoichiometry. The α and β subunits, which make up the bulk of the F_1, are arranged alternately in a hexameric ring; they are homologous to one another and are members of the P-loop NTPase family (Section 9.4). Both bind nucleotides but only the β subunits are catalytically active. Just below the α and β subunits is a central stalk consisting of the γ and ε proteins. The γ subunit includes a long helical coiled coil that extends into the center of the $\alpha_3\beta_3$ hexamer. *The γ subunit breaks the symmetry of the $\alpha_3\beta_3$ hexamer: each of the β subunits is distinct by virtue of its interaction with a different face of γ.* Distinguishing the three β subunits is crucial for understanding the mechanism of ATP synthesis.

The F_0 subunit is a hydrophobic segment that spans the inner mitochondrial membrane. *F_0 contains the proton channel of the complex.* This channel consists of a ring comprising from 8 to 14 **c** subunits that are embedded in the membrane. A single **a** subunit binds

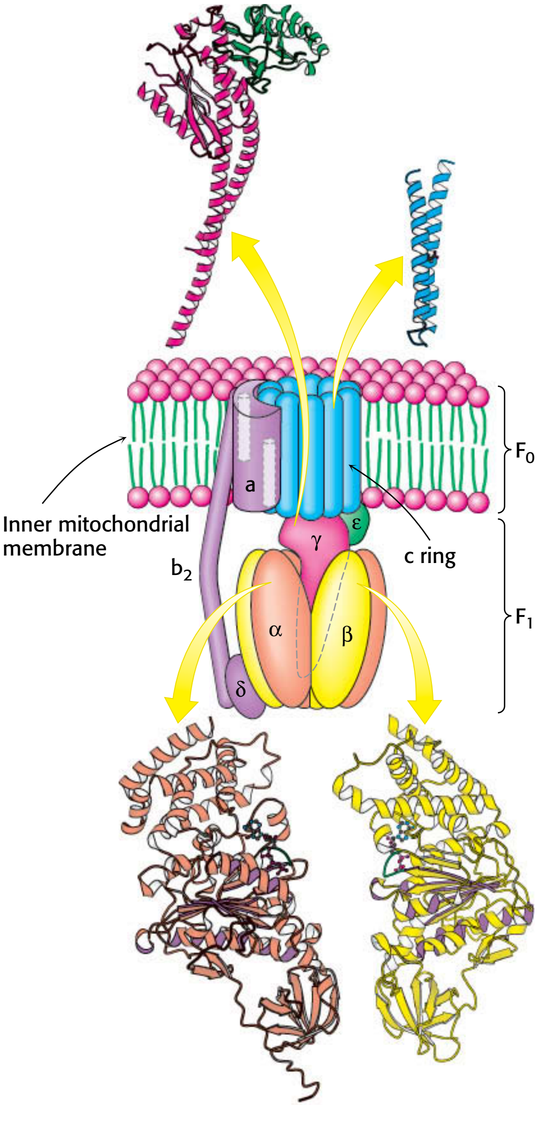


FIGURE 18.24 Structure of ATP synthase. A schematic structure is shown along with representations of the components for which structures have been determined to high resolution. The P-loop NTPase domains of the α and β subunits are indicated by purple shading. *Notice* that part of the enzyme complex is embedded in the inner mitochondrial membrane, whereas the remainder resides in the matrix. [Drawn from 1E79.pdb and 1C0V.pdb.]

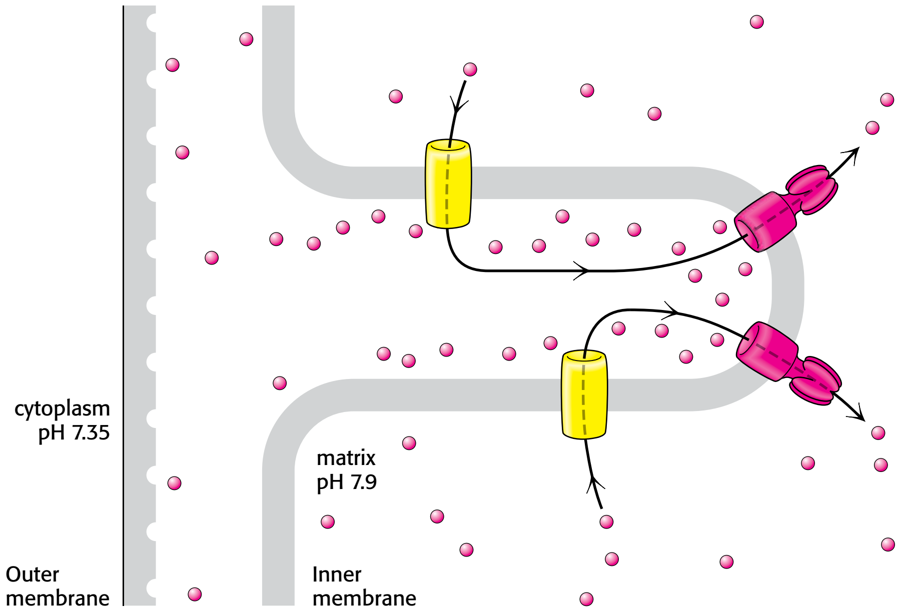


FIGURE 18.25 ATPase assists in the formation of cristae. Oligomers of ATP synthase dimers facilitate cristae formation, creating an area where the protons (red balls) are concentrated and have ready access to the F_o portion of the ATP synthase. The electron transport chain is represented by the yellow cylinders embedded in the inner mitochondrial membrane. [Information from K. M. Davies et al., *Proc. Natl. Acad. Sci. U.S.A.* 108: 14121–14126, 2011.]

to the outside of the ring. The F_0 and F_1 subunits are connected in two ways: by the central $\gamma\varepsilon$ stalk and by an exterior column. The exterior column consists of one **a** subunit, two **b** subunits, and the δ subunit.

ATP synthases associate with one another to form dimers, which then associate to form large oligomers of dimers. This association stabilizes the individual enzymes to the rotational forces required for catalysis and facilitates the curvature of the inner mitochondrial membrane. The formation of the cristae allows the proton pumps of the electron transport chain to localize the proton gradient in the vicinity of the synthases, which are located at the tips of the cristae, thereby enhancing the efficiency of ATP synthesis (Figure 18.25).

Proton flow through ATP synthase leads to the release of tightly bound ATP: The binding-change mechanism

ATP synthase catalyzes the formation of ATP from ADP and orthophosphate.

$$ADP^{3-} + HPO_4^{2-} + H^+ \rightleftharpoons ATP^{4-} + H_2O$$

The actual substrates are ADP and ATP complexed with Mg^{2+}, as in all known phosphoryl-transfer reactions with these nucleotides. A terminal oxygen atom of ADP attacks the phosphorus atom of P_i to form a pentacovalent intermediate, which then dissociates into ATP and H_2O (Figure 18.26).

How does the flow of protons drive the synthesis of ATP? Isotopic-exchange experiments unexpectedly revealed that *enzyme-bound ATP forms readily in the absence of a proton-motive force.* When ADP and P_i were added to ATP synthase in $H_2{}^{18}O$, ^{18}O became incorporated into P_i through the

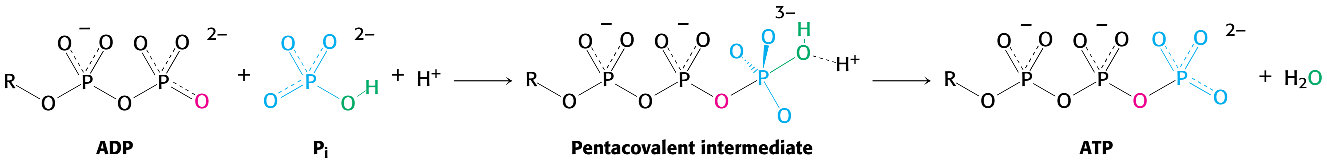


FIGURE 18.26 ATP-synthesis mechanism. One of the oxygen atoms of ADP attacks the phosphorus atom of P_i to form a pentacovalent intermediate, which then forms ATP and releases a molecule of H_2O.

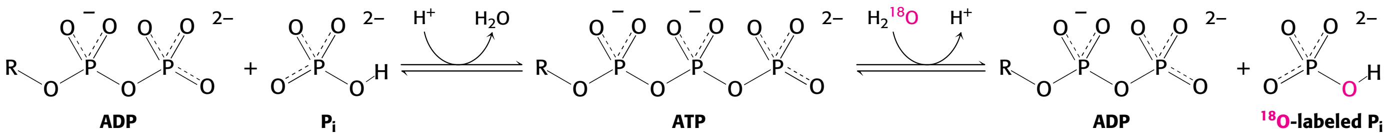


FIGURE 18.27 ATP forms without a proton-motive force but is not released. The results of isotopic-exchange experiments indicate that enzyme-bound ATP is formed from ADP and P_i in the absence of a proton-motive force.

synthesis of ATP and its subsequent hydrolysis (Figure 18.27). The rate of incorporation of ^{18}O into P_i showed that about equal amounts of bound ATP and ADP are in equilibrium at the catalytic site, even in the absence of a proton gradient. However, ATP does not leave the catalytic site unless protons flow through the enzyme. Thus, *the role of the proton gradient is not to form ATP but to release it from the synthase.*

The fact that three β subunits are components of the F_1 moiety of the ATPase means that there are three active sites on the enzyme, each performing one of three different functions at any instant. The proton-motive force causes the three active sites to sequentially change functions as protons flow through the membrane-embedded component of the enzyme. Indeed, we can think of the enzyme as consisting of a moving part and a stationary part: (1) the moving unit, or *rotor,* consists of the **c** ring and the γε stalk and (2) the stationary unit, or *stator,* is composed of the remainder of the molecule.

How do the three active sites of ATP synthase respond to the flow of protons? A number of experimental observations suggested a *binding-change mechanism* for proton-driven ATP synthesis. This proposal states that a β subunit can perform each of three sequential steps in the synthesis of ATP by changing conformation. These steps are (1) ADP and P_i binding, (2) ATP synthesis, and (3) ATP release. As already noted, interactions with the γ subunit make the three β subunits structurally distinct (Figure 18.28). At any given moment, one β subunit will be in the L, or loose, conformation. This conformation binds ADP and P_i. A second subunit will be in the T, or tight, conformation. This conformation binds ATP with great avidity, so much so that it will convert bound ADP and P_i into ATP. Both the T and L conformations are sufficiently constrained that they cannot release bound nucleotides. The final subunit will be in the O, or open, form. This form has a more open conformation and can bind or release adenine nucleotides.

The rotation of the γ subunit drives the interconversion of these three forms (Figure 18.29). ADP and P_i bound in the subunit in the T form are transiently combining to form ATP. Suppose that the γ subunit is rotated by

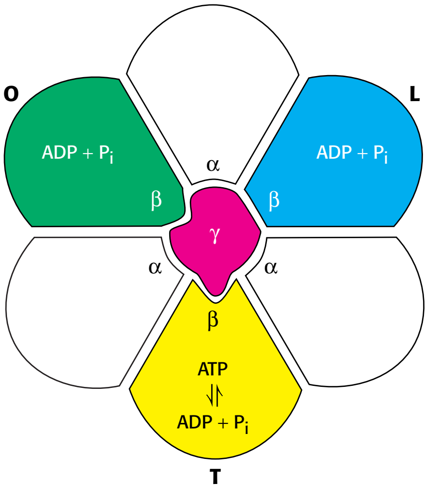


FIGURE 18.28 ATP synthase nucleotide-binding sites are not equivalent. The γ subunit passes through the center of the $\alpha_3\beta_3$ hexamer and makes the nucleotide-binding sites in the β subunits distinct from one another. The β subunits are colored to distinguish them from one another.

Progressive alteration of the forms of the three active sites of ATP synthase

Subunit 1	L → T → O → L → T → O......
Subunit 2	O → L → T → O → L → T......
Subunit 3	T → O → L → T → O → L......

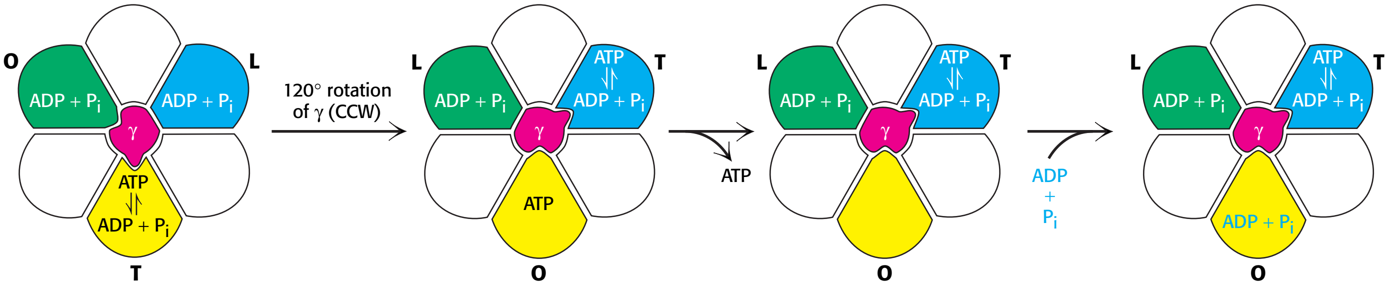


FIGURE 18.29 Binding-change mechanism for ATP synthase. The rotation of the γ subunit interconverts the three β subunits. The subunit in the T (tight) form interconverts ADP and P_i and ATP but does not allow ATP to be released. When the γ subunit is rotated by 120 degrees in a counterclockwise (CCW) direction, the T-form subunit is converted into the O form, allowing ATP release. ADP and P_i can then bind to the O-form subunit. An additional 120-degree rotation (not shown) traps these substrates in an L-form subunit.

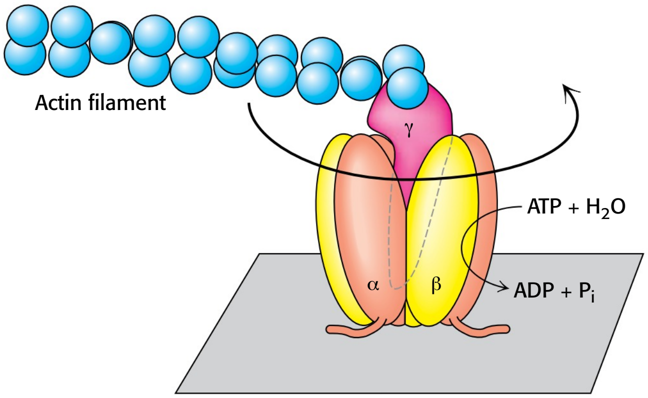


FIGURE 18.30 Direct observation of ATP-driven rotation in ATP synthase. The $\alpha_3\beta_3$ hexamer of ATP synthase is fixed to a surface, with the γ subunit projecting upward and linked to a fluorescently labeled actin filament. The addition and subsequent hydrolysis of ATP result in the counterclockwise rotation of the γ subunit, which can be directly seen under a fluorescence microscope.

120 degrees in a counterclockwise direction (as viewed from the top). This rotation converts the T-form site into an O-form site with the nucleotide bound as ATP. Concomitantly, the L-form site is converted into a T-form site, enabling the transformation of an additional ADP and P_i into ATP. The ATP in the O-form site can now depart from the enzyme to be replaced by ADP and P_i. An additional 120-degree rotation converts this O-form site into an L-form site, trapping these substrates. Each subunit progresses from the T to the O to the L form with no two subunits ever present in the same conformational form. This mechanism suggests that ATP can be synthesized and released by driving the rotation of the γ subunit in the appropriate direction.

Rotational catalysis is the world's smallest molecular motor

Is it possible to observe the proposed rotation directly? Elegant experiments, using single-molecule techniques (Section 8.6), have demonstrated the rotation through the use of a simple experimental system consisting solely of cloned $\alpha_3\beta_3\gamma$ subunits (Figure 18.30). The β subunits were engineered to contain amino-terminal polyhistidine tags, which have a high affinity for nickel ions (Section 3.1). This property of the tags allowed the $\alpha_3\beta_3$ assembly to be immobilized on a glass surface that had been coated with nickel ions. The γ subunit was linked to a fluorescently labeled actin filament to provide a long segment that could be observed under a fluorescence microscope. Remarkably, the addition of ATP caused the actin filament to rotate unidirectionally in a counterclockwise direction. *The γ subunit was rotating, driven by the hydrolysis of ATP.* Thus, the catalytic activity of an individual molecule could be observed. The counterclockwise rotation is consistent with the predicted mechanism for hydrolysis because the molecule was viewed from below relative to the view shown in Figure 18.30.

More-detailed analysis in the presence of lower concentrations of ATP revealed that the γ subunit rotates in 120-degree increments. Each increment corresponds to the hydrolysis of a single ATP molecule. In addition, from the results obtained by varying the length of the actin filament and measuring the rate of rotation, the enzyme appears to operate near 100% efficiency; that is, essentially all of the energy released by ATP hydrolysis is converted into rotational motion.

Proton flow around the c ring powers ATP synthesis

The direct observation of rotary motion of the γ subunit is strong evidence for the rotational mechanism for ATP synthesis. The last remaining question is: How does proton flow through F_0 drive the rotation of the γ subunit? The mechanism depends on the structures of the **a** and **c** subunits of F_0 (Figure 18.31). The stationary **a** subunit directly abuts the membrane-spanning ring formed by 8 to 14 **c** subunits. The **a** subunit includes two

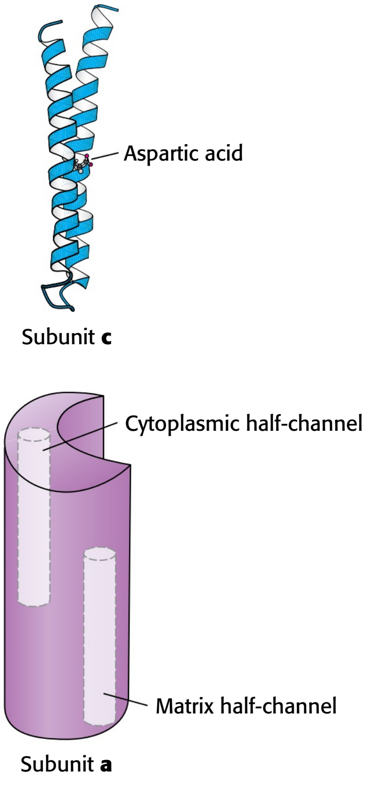
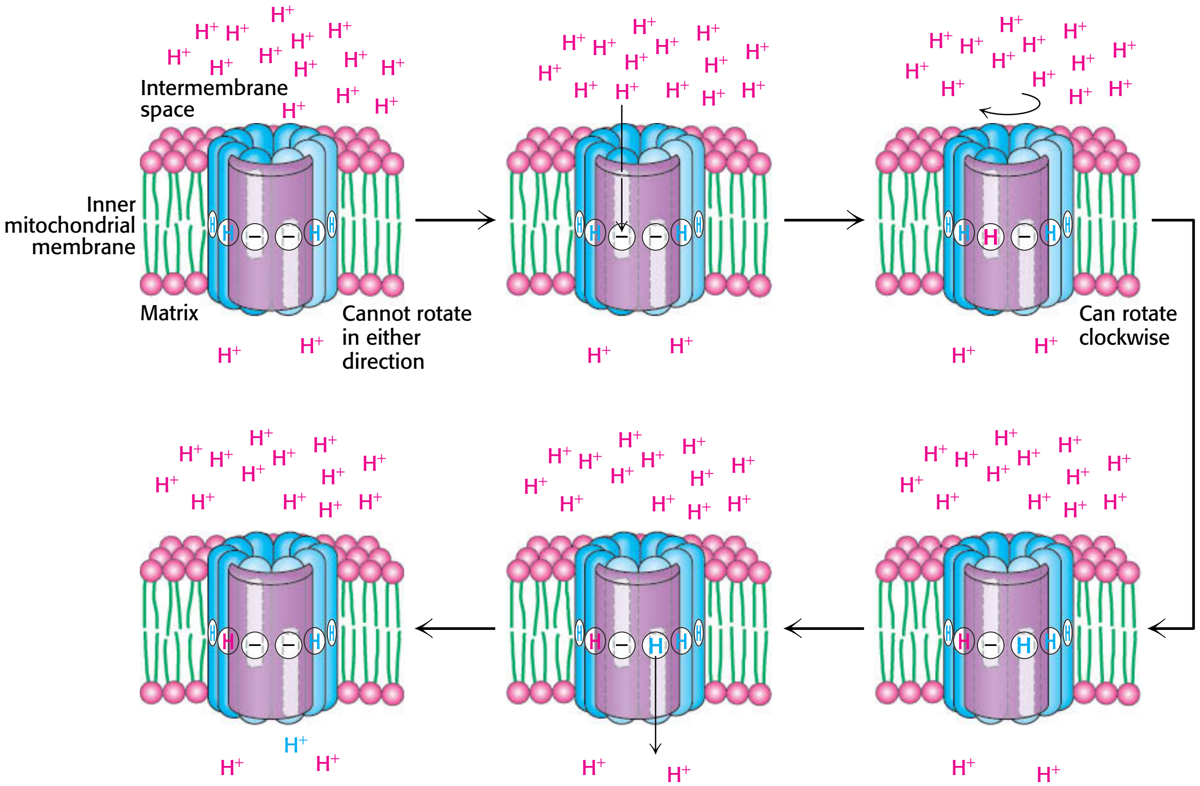


FIGURE 18.31 Components of the proton-conducting unit of ATP synthase. The **c** subunit consists of two α helices that span the membrane. In *E. coli*, an aspartic acid residue in one of the helices lies on the center of the membrane. The structure of the **a** subunit has not yet been directly observed, but it appears to include two half-channels that allow protons to enter and pass partway but not completely through the membrane.



Intermembrane space
Inner mitochondrial membrane
Matrix
Cannot rotate in either direction
Can rotate clockwise
H+
H

FIGURE 18.32 **Proton motion across the membrane drives rotation of the c ring.** A proton enters from the intermembrane space into the cytoplasmic half-channel to neutralize the charge on an aspartate residue in a **c** subunit. With this charge neutralized, the **c** ring can rotate clockwise by one **c** subunit, moving an aspartic acid residue out of the membrane into the matrix half-channel. This proton can move into the matrix, resetting the system to its initial state.

hydrophilic half-channels that do not span the membrane (Figure 18.31). Thus, protons can pass into either of these channels, but they cannot move completely across the membrane. The **a** subunit is positioned such that each half-channel directly interacts with one **c** subunit.

The structure of the **c** subunit was determined both by NMR methods and by x-ray crystallography. Each polypeptide chain forms a pair of α helices that span the membrane. A glutamic acid (or aspartic acid) residue is found in the middle of one of the helices. If the glutamate is charged (unprotonated), the **c** subunit will not move into the membrane. The key to proton movement across the membrane is that, in a proton-rich environment, such as the cytoplasmic side of the inner mitochondrial membrane, a proton will enter a channel and bind the glutamate residue, while the glutamic acid in a proton-poor environment of the other half-channel will release a proton (Figure 18.32). The **c** subunit with the bound proton then moves into the membrane as the ring rotates by one **c** subunit. This rotation brings the newly deprotonated **c** subunit from the matrix half-channel to the proton rich cytoplasmic half-channel, where it can bind a proton. *The movement of protons through the half-channels from the high proton concentration of the cytoplasm to the low proton concentration of the matrix powers the rotation of the* **c** *ring.* The **a** unit remains stationary as the **c** ring rotates. Each proton that enters the cytoplasmic half-channel of the **a** unit moves through the membrane by riding around on the rotating **c** ring to exit through the matrix half-channel into the proton-poor environment of the matrix (Figure 18.33).

How does the rotation of the **c** ring lead to the synthesis of ATP? The **c** ring is tightly linked to the γ and ε subunits. Thus, as the **c** ring turns, the

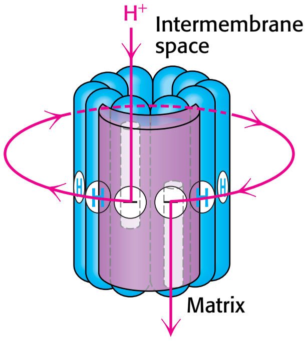


FIGURE 18.33 **Proton path through the membrane.** Each proton enters the cytoplasmic half-channel, follows a complete rotation of the **c** ring, and exits through the other half-channel into the matrix.

γ and ε subunits are turned inside the $\alpha_3\beta_3$ hexamer unit of F_1. The rotation of the γ subunit in turn promotes the synthesis of ATP through the binding-change mechanism. The exterior column formed by the two **b** chains and the δ subunit prevents the $\alpha_3\beta_3$ hexamer from rotating in sympathy with the **c** ring. Recall that the number of **c** subunits in the **c** ring appears to range between 8 and 14. This number is significant because it determines the number of protons that must be transported to generate a molecule of ATP. Each 360-degree rotation of the γ subunit leads to the synthesis and release of three molecules of ATP. Thus, if there are 10 **c** subunits in the ring (as was observed in a crystal structure of yeast mitochondrial ATP synthase), each ATP generated requires the transport of $10/3 = 3.33$ protons. Recent evidence shows that the **c** rings of all vertebrates are composed of 8 subunits, making vertebrate ATP synthase the most-efficient ATP synthase known, with the transport of only 2.7 protons required for ATP synthesis. For simplicity, we will assume that three protons must flow into the matrix for each ATP formed, but we must keep in mind that the true value may differ. As we will see, the electrons from NADH pump enough protons to generate 2.5 molecules of ATP, whereas those from $FADH_2$ yield 1.5 molecules of ATP.

A little goes a long way

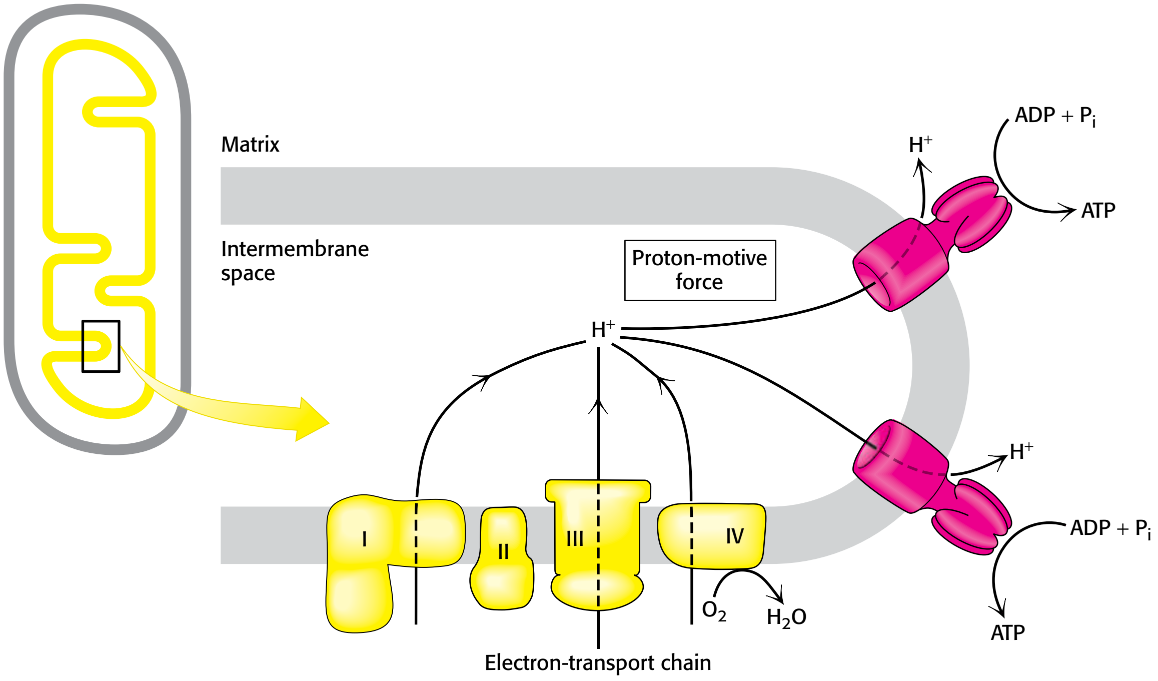
Despite the various molecular machinations and the vast numbers of ATPs synthesized and protons pumped, a resting human being requires surprisingly little power. Approximately 116 watts, the energy output of a typical light bulb, provides enough energy to sustain a resting person.

Let us return for a moment to the example with which we began this chapter. If a resting human being requires 85 kg of ATP per day for bodily functions, then 3.3×10^{25} protons must flow through the ATP synthase per day, or 3.3×10^{21} protons per second. Figure 18.34 summarizes the process of oxidative phosphorylation.

ATP synthase and G proteins have several common features

The α and β subunits of ATP synthase are members of the P-loop NTPase family of proteins. In Chapter 14, we learned that the signaling properties of other members of this family, the G proteins, depend on their ability to bind nucleoside triphosphates and diphosphates with great

FIGURE 18.34 **Overview of oxidative phosphorylation.** The electron-transport chain generates a proton gradient, which is used to synthesize ATP.



tenacity. They do not exchange nucleotides unless they are stimulated to do so by interaction with other proteins. The binding-change mechanism of ATP synthase is a variation on this theme. The P-loop regions of the β subunits will bind either ADP or ATP (or release ATP), depending on which of three different faces of the γ subunit they interact with. The conformational changes take place in an orderly way, driven by the rotation of the γ subunit.

18.5 Many Shuttles Allow Movement Across Mitochondrial Membranes

The inner mitochondrial membrane must be impermeable to most molecules, yet much exchange has to take place between the cytoplasm and the mitochondria. This exchange is mediated by an array of membrane-spanning transporter proteins (Section 13.4).

Electrons from cytoplasmic NADH enter mitochondria by shuttles

One function of the respiratory chain is to regenerate NAD^+ for use in glycolysis. NADH generated by the citric acid cycle and fatty acid oxidation is already in the mitochondrial matrix, but how is cytoplasmic NADH reoxidized to NAD^+ under aerobic conditions? NADH cannot simply pass into mitochondria for oxidation by the respiratory chain, because the inner mitochondrial membrane is impermeable to NADH and NAD^+. The solution is that *electrons from NADH,* rather than NADH itself, are carried across the mitochondrial membrane. One of several means of introducing electrons from NADH into the electron-transport chain is the *glycerol 3-phosphate shuttle* (Figure 18.35). The first step in this shuttle is the transfer of a pair of electrons from NADH to dihydroxyacetone phosphate, a glycolytic intermediate, to form glycerol 3-phosphate. This reaction is catalyzed by a glycerol 3-phosphate dehydrogenase in the cytoplasm. Glycerol 3-phosphate is reoxidized to dihydroxyacetone phosphate on the outer surface of the inner mitochondrial membrane by a membrane-bound isozyme of glycerol 3-phosphate dehydrogenase. An electron pair from glycerol 3-phosphate is

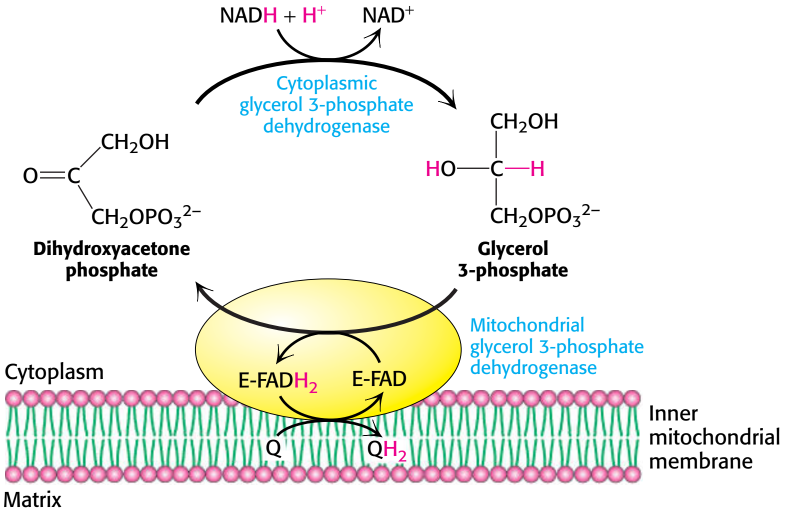
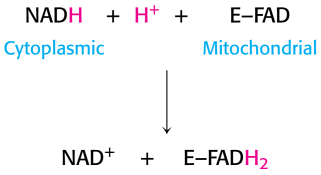

NADH + H+
NAD+
Cytoplasmic glycerol 3-phosphate dehydrogenase
CH2OH
O═C
CH2OPO3 2−
Dihydroxyacetone phosphate
CH2OH
HO—C—H
CH2OPO3 2−
Glycerol 3-phosphate
Mitochondrial glycerol 3-phosphate dehydrogenase
Cytoplasm
E-FADH2
E-FAD
Q
QH2
Inner mitochondrial membrane
Matrix

FIGURE 18.35 Glycerol 3-phosphate shuttle. Electrons from NADH can enter the mitochondrial electron-transport chain by being used to reduce dihydroxyacetone phosphate to glycerol 3-phosphate. Glycerol 3-phosphate is reoxidized by electron transfer to an FAD prosthetic group in a membrane-bound glycerol 3-phosphate dehydrogenase. Subsequent electron transfer to Q to form QH_2 allows these electrons to enter the electron-transport chain.

transferred to an FAD prosthetic group in this enzyme to form $FADH_2$. This reaction also regenerates dihydroxyacetone phosphate.

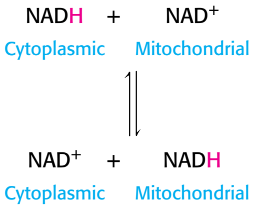
The reduced flavin transfers its electrons to the electron carrier Q, which then enters the respiratory chain as QH_2. *When cytoplasmic NADH transported by the glycerol 3-phosphate shuttle is oxidized by the respiratory chain, 1.5 rather than 2.5 molecules of ATP are formed.* The yield is lower because FAD rather than NAD^+ is the electron acceptor in mitochondrial glycerol 3-phosphate dehydrogenase. The use of FAD enables electrons from cytoplasmic NADH to be transported into mitochondria against an NADH concentration gradient. The price of this transport is one molecule of ATP per two electrons. This glycerol 3-phosphate shuttle is especially prominent in muscle and enables it to sustain a very high rate of oxidative phosphorylation. Indeed, some insects lack lactate dehydrogenase and are completely dependent on the glycerol 3-phosphate shuttle for the regeneration of cytoplasmic NAD^+.



$$\underset{\text{Cytoplasmic}}{\text{NADH}} + \text{H}^+ + \underset{\text{Mitochondrial}}{\text{E–FAD}} \longrightarrow \underset{\text{Cytoplasmic}}{\text{NAD}^+} + \underset{\text{Mitochondrial}}{\text{E–FADH}_2}$$

Glycerol 3-phosphate shuttle

In the heart and liver, electrons from cytoplasmic NADH are brought into mitochondria by the *malate–aspartate shuttle,* which is mediated by two membrane carriers and four enzymes (Figure 18.36). Electrons are transferred from NADH in the cytoplasm to oxaloacetate, forming malate, which traverses the inner mitochondrial membrane in exchange for α-ketoglutarate and is then reoxidized by NAD^+ in the matrix to form NADH in a reaction catalyzed by the citric acid cycle enzyme malate dehydrogenase. The resulting oxaloacetate does not readily cross the inner mitochondrial membrane and so a transamination reaction (Section 23.3) is needed to form aspartate, which can be transported to the cytoplasmic side in exchange for glutamate. Glutamate donates an amino group to oxaloacetate, forming aspartate and α-ketoglutarate. In the cytoplasm, aspartate is then deaminated to form oxaloacetate and the cycle is restarted.



$$\underset{\text{Cytoplasmic}}{\text{NADH}} + \underset{\text{Mitochondrial}}{\text{NAD}^+} \rightleftharpoons \underset{\text{Cytoplasmic}}{\text{NAD}^+} + \underset{\text{Mitochondrial}}{\text{NADH}}$$

Malate–aspartate shuttle

The entry of ADP into mitochondria is coupled to the exit of ATP by ATP-ADP translocase

The major function of oxidative phosphorylation is to generate ATP from ADP. ATP and ADP do not diffuse freely across the inner mitochondrial membrane. How are these highly charged molecules moved across the inner membrane into the cytoplasm? A specific transport protein, *ATP-ADP translocase,* enables these molecules to transverse this permeability barrier.

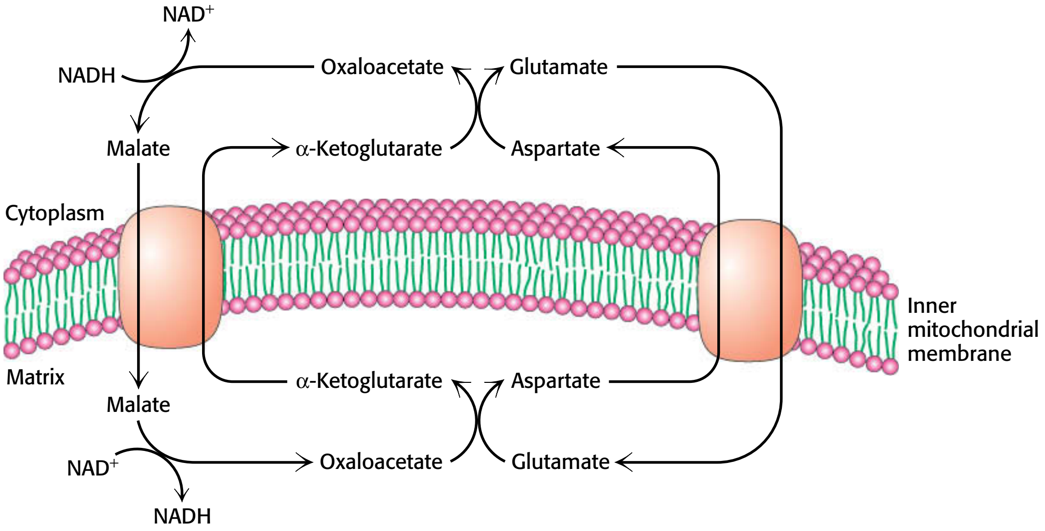


FIGURE 18.36 Malate–aspartate shuttle.

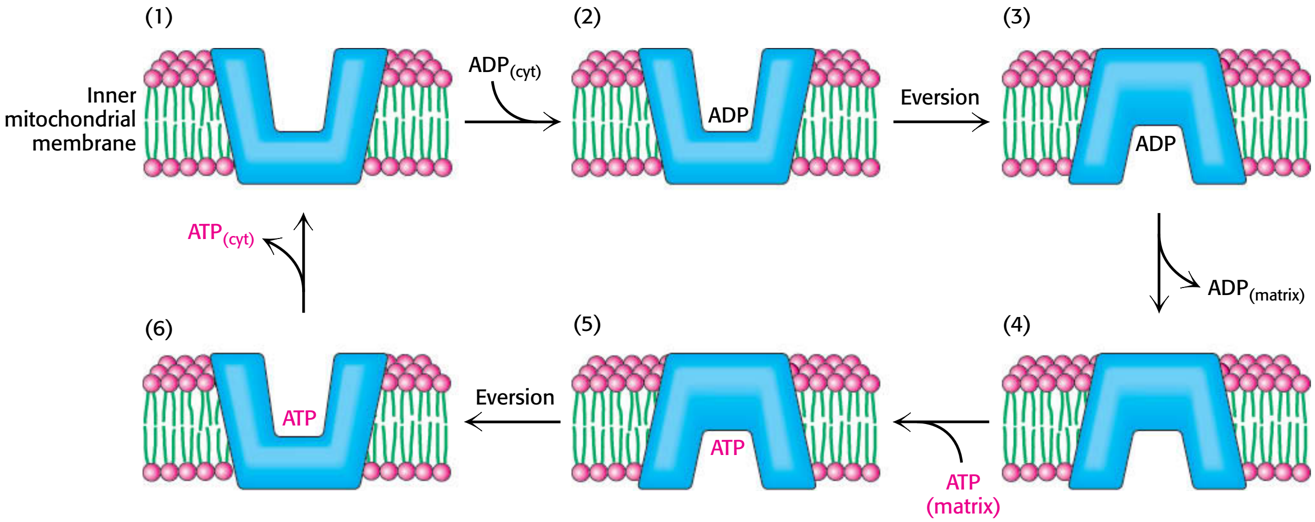


FIGURE 18.37 Mechanism of mitochondrial ATP-ADP translocase. The translocase catalyzes the coupled entry of ADP into the matrix and the exit of ATP from it. The binding of ADP (1) from the cytoplasm favors eversion of the transporter (2) to release ADP into the matrix (3). Subsequent binding of ATP from the matrix to the everted form (4) favors eversion back to the original conformation (5), releasing ATP into the cytoplasm (6).

Most important, the flows of ATP and ADP are coupled. *ADP enters the mitochondrial matrix only if ATP exits, and vice versa.* This process is carried out by the translocase, an antiporter:

$$\mathrm{ADP^{3-}}_{\mathrm{cytoplasm}} + \mathrm{ATP^{4-}}_{\mathrm{matrix}} \rightarrow \mathrm{ADP^{3-}}_{\mathrm{matrix}} + \mathrm{ATP^{4-}}_{\mathrm{cytoplasm}}$$

ATP-ADP translocase is highly abundant, constituting about 15% of the protein in the inner mitochondrial membrane. The abundance is a manifestation of the fact that human beings exchange the equivalent of their weight in ATP each day. Despite the fact that mitochondria are the sites of ATP synthesis, both the cytoplasm and the nucleus have more ATP than the mitochondria, a testament to the efficiency of transport.

The 30-kDa translocase contains a single nucleotide-binding site that alternately faces the matrix and the cytoplasmic sides of the membrane (Figure 18.37). ATP and ADP bind to the translocase without $\mathrm{Mg^{2+}}$, and ATP has one more negative charge than that of ADP. Thus, in an actively respiring mitochondrion with a positive membrane potential, ATP transport out of the mitochondrial matrix and ADP transport into the matrix are favored. This ATP-ADP exchange is energetically expensive; about a quarter of the proton-motive force generated by the respiratory chain is consumed by this exchange process. The inhibition of the translocase leads to the subsequent inhibition of cellular respiration as well (p. 557).

Mitochondrial transporters for metabolites have a common tripartite structure

Examination of the amino acid sequence of the ATP-ADP translocase revealed that this protein consists of three tandem repeats of a 100-amino-acid module, each of which appears to have two transmembrane segments. This tripartite structure has recently been confirmed by the determination of the three-dimensional structure of this transporter (Figure 18.38). The transmembrane helices form a tepee-like structure with the nucleotide-binding site (marked by a bound inhibitor) lying in the center. Each of the three repeats adopts a similar structure.

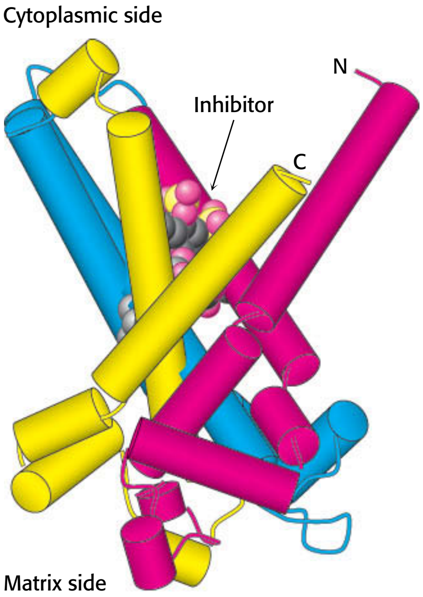


FIGURE 18.38 Structure of mitochondrial transporters. The structure of the ATP-ADP translocase is shown. *Notice* that this structure comprises three similar units (shown in red, blue, and yellow) that come together to form a binding site, here occupied by atractyloside, an inhibitor of this transporter. Other members of the mitochondrial transporter family adopt similar tripartite structures. [Drawn from 1OKC.pdb.]

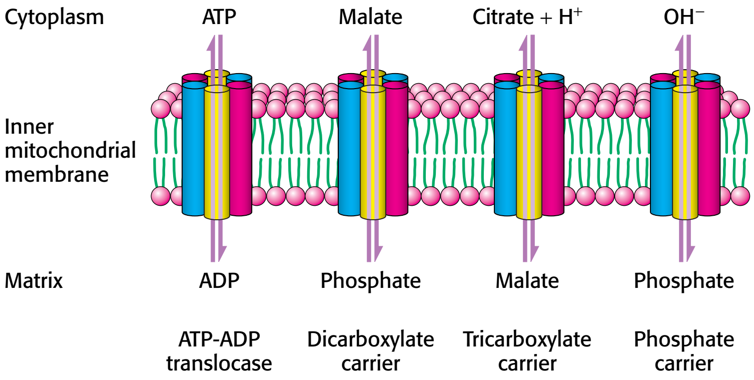


FIGURE 18.39 Mitochondrial transporters. Transporters (also called carriers) are transmembrane proteins that carry specific ions and charged metabolites across the inner mitochondrial membrane.

ATP-ADP translocase is but one of many mitochondrial transporters for ions and charged metabolites (Figure 18.39). The *phosphate carrier,* which works in concert with ATP-ADP translocase, mediates the electroneutral exchange of $H_2PO_4^-$ for OH^-. The combined action of these two transporters leads to the exchange of cytoplasmic ADP and P_i for matrix ATP at the cost of the influx of one H^+ (owing to the transport of one OH^- out of the matrix). These two transporters, which provide ATP synthase with its substrates, are associated with the synthase to form a large complex called the *ATP synthasome.*

Other homologous carriers also are present in the inner mitochondrial membrane. The dicarboxylate carrier enables malate, succinate, and fumarate to be exported from the mitochondrial matrix in exchange for P_i. The tricarboxylate carrier exchanges citrate and H^+ for malate. In all, more than 40 such carriers are encoded in the human genome. A different type of carrier is responsible for pyruvate transport into the mitochondria. Pyruvate in the cytoplasm enters the mitochondrial membrane by way of a heterodimer composed of two small transmembrane proteins.

18.6 The Regulation of Cellular Respiration Is Governed Primarily by the Need for ATP

Because ATP is the end product of cellular respiration, the ATP needs of the cell are the ultimate determinant of the rate of respiratory pathways and their components.

The complete oxidation of glucose yields about 30 molecules of ATP

We can now estimate how many molecules of ATP are formed when glucose is completely oxidized to CO_2. The number of ATP molecules formed in glycolysis and the citric acid cycle is unequivocally known because it is determined by the stoichiometries of chemical reactions. In contrast, the ATP yield of oxidative phosphorylation is less certain because the stoichiometries of proton pumping, ATP synthesis, and metabolite-transport processes need not be an integer or even have fixed values. As stated earlier, the best current estimates for the number of protons pumped out of the matrix by NADH-Q oxidoreductase, Q-cytochrome *c* oxidoreductase, and cytochrome *c* oxidase per electron pair are four, two, and four, respectively. The synthesis of a molecule of ATP is driven by the flow of about three protons through ATP synthase. An additional proton is consumed in transporting ATP from the matrix to the cytoplasm. Hence, about 2.5 molecules of cytoplasmic ATP are generated as a result of the flow of a pair of electrons from NADH to O_2. For electrons that enter at the level of

TABLE 18.4 ATP yield from the complete oxidation of glucose

Reaction sequence	ATP yield per glucose molecule
Glycolysis: Conversion of glucose into pyruvate (in the cytoplasm)	
Phosphorylation of glucose	−1
Phosphorylation of fructose 6-phosphate	−1
Dephosphorylation of 2 molecules of 1,3-BPG	+2
Dephosphorylation of 2 molecules of phosphoenolpyruvate	+2
2 molecules of NADH are formed in the oxidation of 2 molecules of glyceraldehyde 3-phosphate	
Conversion of pyruvate into acetyl CoA (inside mitochondria)	
2 molecules of NADH are formed	
Citric acid cycle (inside mitochondria)	
2 molecules of adenosine triphosphate are formed from 2 molecules of succinyl CoA	+2
6 molecules of NADH are formed in the oxidation of 2 molecules each of isocitrate, α-ketoglutarate, and malate	
2 molecules of $FADH_2$ are formed in the oxidation of 2 molecules of succinate	
Oxidative phosphorylation (inside mitochondria)	
2 molecules of NADH formed in glycolysis; each yields 1.5 molecules of ATP (assuming transport of NADH by the glycerol 3-phosphate shuttle)	+3
2 molecules of NADH formed in the oxidative decarboxylation of pyruvate; each yields 2.5 molecules of ATP	+5
2 molecules of $FADH_2$ formed in the citric acid cycle; each yields 1.5 molecules of ATP	+3
6 molecules of NADH formed in the citric acid cycle; each yields 2.5 molecules of ATP	+15
Net Yield per Molecule of Glucose	+30

Information on the ATP yield of oxidative phosphorylation is from values given in P. C. Hinkle, M. A. Kumar, A. Resetar, and D. L. Harris, *Biochemistry* 30:3576, 1991.

Note: The current value of 30 molecules of ATP per molecule of glucose supersedes the earlier value of 36 molecules of ATP. The stoichiometries of proton pumping, ATP synthesis, and metabolite transport should be regarded as estimates. About 2 more molecules of ATP are formed per molecule of glucose oxidized when the malate–aspartate shuttle rather than the glycerol 3-phosphate shuttle is used.

Q-cytochrome *c* oxidoreductase, such as those from the oxidation of succinate or cytoplasmic NADH transferred by the glycerol-phosphate shuttle, the yield is about 1.5 molecules of ATP per electron pair. Hence, as tallied in Table 18.4, *about 30 molecules of ATP are formed when glucose is completely oxidized to* CO_2. Most of the ATP, 26 of 30 molecules formed, is generated by oxidative phosphorylation. Recall that the anaerobic metabolism of glucose yields only 2 molecules of ATP. One of the effects of endurance exercise, a practice that calls for much ATP for an extended period of time, is to increase the number of mitochondria and blood vessels in muscle and thus increase the extent of ATP generation by oxidative phosphorylation.

The rate of oxidative phosphorylation is determined by the need for ATP

How is the rate of the electron-transport chain controlled? Under most physiological conditions, electron transport is tightly coupled to phosphorylation. *Electrons do not usually flow through the electron-transport chain to* O_2 *unless ADP is simultaneously phosphorylated to ATP.* When ADP concentration rises, as would be the case in active muscle, the rate of oxidative phosphorylation increases to meet the ATP needs of the muscle. The regulation of the rate of oxidative phosphorylation by the ADP level is called *respiratory control* or *acceptor control.* Experiments on isolated mitochondria demonstrate the

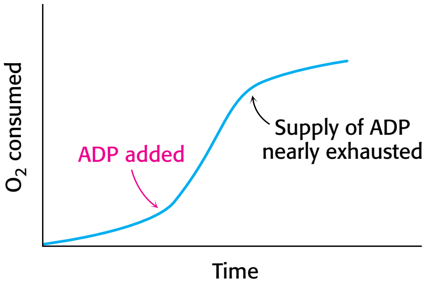


FIGURE 18.40 Respiratory control. Electrons are transferred to O_2 only if ADP is concomitantly phosphorylated to ATP.

importance of ADP level (Figure 18.40). The rate of oxygen consumption by mitochondria increases markedly when ADP is added and then returns to its initial value when the added ADP has been converted into ATP.

The level of ADP likewise affects the rate of the citric acid cycle. At low concentrations of ADP, as in a resting muscle, NADH and $FADH_2$ are not consumed by the electron-transport chain. The citric acid cycle slows because there is less NAD^+ and FAD to feed the cycle. As the ADP level rises and oxidative phosphorylation speeds up, NADH and $FADH_2$ are oxidized, and the citric acid cycle becomes more active. *Electrons do not flow from fuel molecules to O_2 unless ATP needs to be synthesized.* We see here another example of the regulatory significance of the energy charge (Figure 18.41).

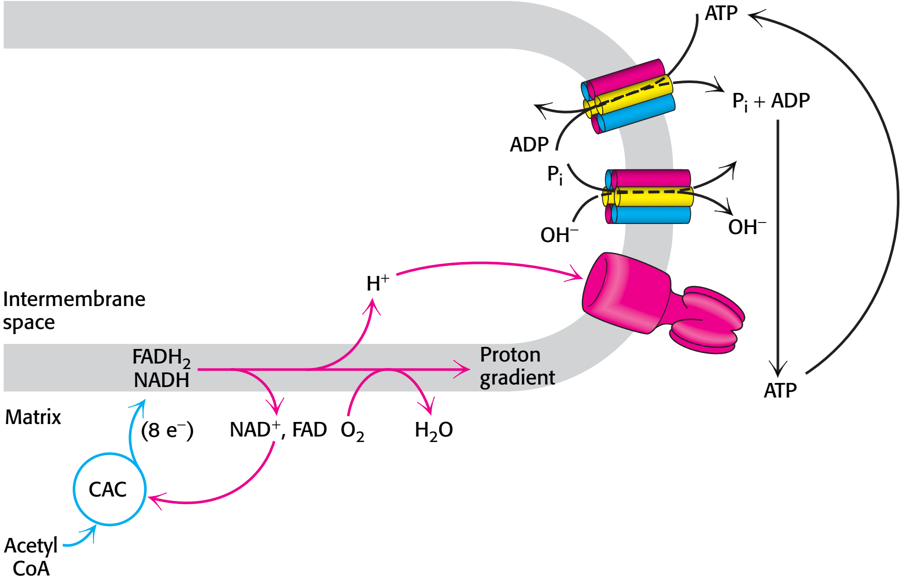


FIGURE 18.41 Energy charge regulates the use of fuels. The synthesis of ATP from ADP and P_i controls the flow of electrons from NADH and $FADH_2$ to oxygen. The availability of NAD^+ and FAD in turn control the rate of the citric acid cycle (CAC).

ATP synthase can be regulated

Mitochondria contain an evolutionarily conserved protein, *inhibitory factor 1* (IF1), that specifically inhibits the potential hydrolytic activity of the F_0F_1 ATP synthase. What is the function of IF1? Consider a circumstance where tissues may be deprived of oxygen (ischemia). Without oxygen as the electron acceptor, the electron transport chain will be unable to generate the proton-motive force. The ATP in the mitochondria would be hydrolyzed by the synthase, working in reverse (Problem 18.45). The role of IF1 is to prevent the wasteful hydrolysis of ATP by inhibiting the hydrolytic activity of the synthase.

IF1 is overexpressed in many types of cancer. This over-expression plays a role in the induction of the Warburg effect, the switch from oxidative phosphorylation to aerobic glycolysis as the principle means for ATP synthesis (Section 16.2).

Regulated uncoupling leads to the generation of heat

Some organisms possess the ability to uncouple oxidative phosphorylation from ATP synthesis to generate heat. Such uncoupling is a means to maintain body temperature in hibernating animals, in some newborn animals (including human beings), and in many adult mammals, especially those adapted to cold. The skunk cabbage uses an analogous mechanism to heat its floral spikes in early spring, increasing the evaporation of odoriferous molecules that attract insects to fertilize its flowers. In animals, the uncoupling is in *brown adipose tissue* (BAT), which is specialized tissue for the process of *nonshivering* *thermogenesis*. In contrast, *white adipose tissue* (WAT), which constitutes the bulk of adipose tissue, plays no role in thermogenesis but serves as an energy source and an endocrine gland (Chapters 26 and 27).

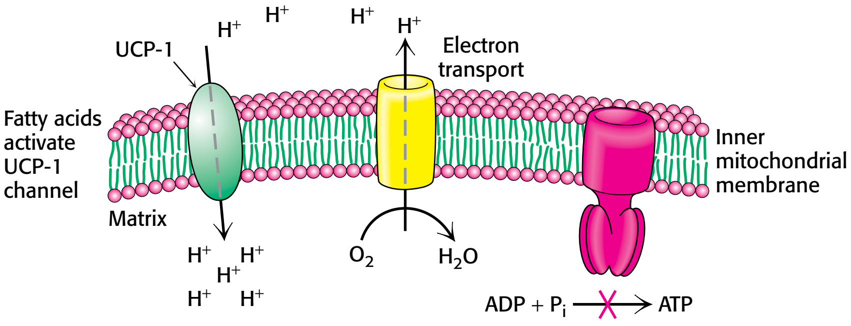


FIGURE 18.42 Action of an uncoupling protein. Uncoupling protein (UCP-1) generates heat by permitting the influx of protons into the mitochondria without the synthesis of ATP.

Brown adipose tissue is very rich in mitochondria, often called *brown fat mitochondria*. The tissue appears brown from the combination of the greenish-colored cytochromes in the numerous mitochondria and the red hemoglobin present in the extensive blood supply, which helps to carry the heat through the body. The inner mitochondrial membrane of these mitochondria contains a large amount of *uncoupling protein* (UCP-1), or *thermogenin,* a dimer of 33-kDa subunits that resembles ATP-ADP translocase. UCP-1 transports protons from the cytoplasm to the matrix with the assistance of fatty acids. In essence, *UCP-1 generates heat by short-circuiting the mitochondrial proton battery.* The energy of the proton gradient, normally captured as ATP, is released as heat as the protons flow through UCP-1 to the mitochondrial matrix. This dissipative proton pathway is activated when the core body temperature begins to fall. In response to a temperature drop, α-adrenergic hormones stimulate the release of free fatty acids from triacylglycerols stored in cytoplasmic lipid granules (Section 22.2) (Figure 18.42). Long chain fatty acids bind to the cytoplasmic face of UCP-1, and the carboxyl group binds a proton. This causes a structural change in UCP-1 so that the protonated carboxyl now faces the proton-poor environment of the matrix, and the proton is released. Proton release resets UCP-1 to the initial state.

Until recently, adult humans were believed to lack brown fat tissue. However, new studies have established that adults, especially females, have brown adipose tissue in the neck and upper chest regions that is activated by cold (Figure 18.43). Obesity leads to a decrease in brown adipose tissue.

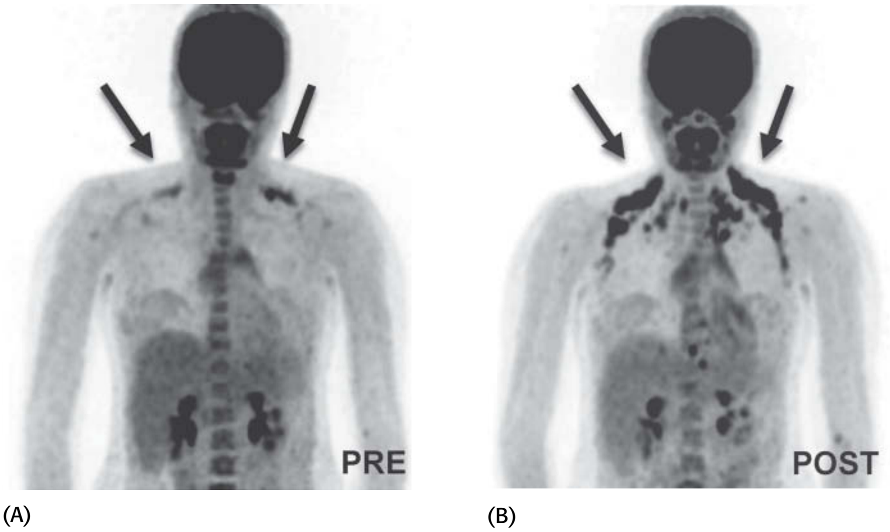


FIGURE 18.43 Brown adipose tissue is revealed on exposure to cold. The results of PET–CT scanning show the uptake and distribution of ^{18}F-fluorodeoxyglucose (^{18}F-FDG) in adipose tissue. The patterns of ^{18}F-FDG uptake in the same subject are dramatically different under thermoneutral conditions (A) and after exposure to cold (B). [Republished with permission of American Society for Clinical Investigation from J Clin Invest. 2013;123(8):3395–3403. doi:10.1172/JCI68993. Permission conveyed through Copyright Clearance Center, Inc.]

We can witness the effects of a lack of nonshivering thermogenesis by examining pig behavior. Pigs are unusual mammals in that they have large litters and are the only ungulates (hoofed animals) that build nests for birth. These behavioral characteristics appear to be the result of a biochemical deficiency. Pigs lack UCP-1 and, hence, brown fat. Piglets must rely on other means of thermogenesis, such as nesting, large litter size, and shivering.

In addition to UCP-1, two other uncoupling proteins have been identified. UCP-2, which is 56% identical in sequence with UCP-1, is found in a wide variety of tissues. UCP-3 (57% identical with UCP-1 and 73% identical with UCP-2) is localized to skeletal muscle and brown fat. This family of uncoupling proteins, especially UCP-2 and UCP-3, may play a role in energy homeostasis. In fact, the genes for UCP-2 and UCP-3 map to regions of the human and mouse chromosomes that have been linked to obesity, supporting the notion that they function as a means of regulating body weight.

Oxidative phosphorylation can be inhibited at many stages

Many potent and lethal poisons exert their effect by inhibiting oxidative phosphorylation at one of a number of different locations (Figure 18.44):

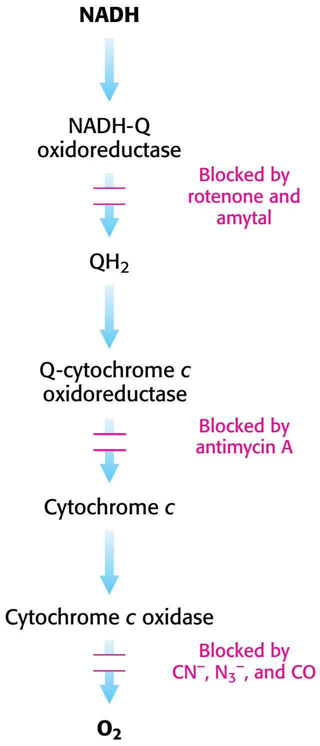
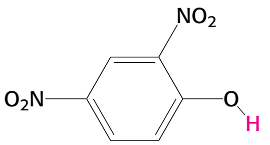

NADH
NADH-Q oxidoreductase
Blocked by rotenone and amytal
QH_2
Q-cytochrome c oxidoreductase
Blocked by antimycin A
Cytochrome c
Cytochrome c oxidase
Blocked by CN^-, N_3^-, and CO
O_2

FIGURE 18.44 Sites of action of some inhibitors of electron transport.

1. *Inhibition of the electron-transport chain. Rotenone,* which is used as a fish and insect poison, and *amytal,* a barbiturate sedative, block electron transfer in NADH-Q oxidoreductase and thereby prevent the utilization of NADH as a substrate. Rotenone, as an electron-transport-chain inhibitor, may play a role, along with genetic susceptibility, in the development of Parkinson disease. In the presence of rotenone and amytal, electron flow resulting from the oxidation of succinate is unimpaired, because these electrons enter through QH_2, beyond the block. *Antimycin A* interferes with electron flow from cytochrome b_H in Q-cytochrome *c* oxidoreductase. Furthermore, electron flow in cytochrome *c* oxidase can be blocked by *cyanide* (CN^-), *azide* (N_3^-), and *carbon monoxide* (CO). Cyanide and azide react with the ferric form of heme a_3, whereas carbon monoxide inhibits the ferrous form. Inhibition of the electron-transport chain also inhibits ATP synthesis because the proton-motive force can no longer be generated.

2. *Inhibition of ATP synthase.* Oligomycin, an antibiotic used as an antifungal agent, and dicyclohexylcarbodiimide (DCC) prevent the influx of protons through ATP synthase by binding to the carboxylate group of the **c** subunits required for proton binding. Modification of only one **c** subunit by DCC is sufficient to inhibit the rotation of the entire **c** ring and hence ATP synthesis. If actively respiring mitochondria are exposed to an inhibitor of ATP synthase, the electron-transport chain ceases to operate. This observation clearly illustrates that electron transport and ATP synthesis are normally tightly coupled.

3. *Uncoupling electron transport from ATP synthesis.* The tight coupling of electron transport and phosphorylation in mitochondria can be uncoupled by 2,4-dinitrophenol (DNP) and certain other acidic aromatic compounds. These substances carry protons across the inner mitochondrial membrane, down their concentration gradient. In the presence of these uncouplers, electron transport from NADH to O_2 proceeds in a normal fashion, but ATP is not formed by mitochondrial ATP synthase, because the proton-motive force across the inner mitochondrial membrane is continuously dissipated. This loss of respiratory control leads to increased oxygen consumption and oxidation of NADH. Indeed, in the accidental ingestion of uncouplers, large amounts of metabolic fuels are consumed, but no energy is captured as ATP. Rather, energy is released as heat. DNP is the active



2,4-Dinitrophenol (DNP)

ingredient in some herbicides and fungicides. Remarkably, some people consume DNP as a weight-loss drug, despite the fact that the FDA banned its use in 1938. There are also reports that Soviet soldiers were given DNP to keep them warm during the long Russian winters. Chemical uncouplers are nonphysiological, unregulated counterparts of uncoupling proteins.

Drugs are being sought that would function as mild uncouplers, uncouplers not as potentially lethal as DNP, for use in treatment of obesity and related pathologies. Xanthohumol, a prenylated chalcone found in hops and beer, shows promise in this regard. Xanthohumol also scavenges free radicals and is used for treatment of certain types of cancers.

4. *Inhibition of ATP export.* ATP-ADP translocase is specifically inhibited by very low concentrations of *atractyloside* (a plant glycoside) or *bongkrekic acid* (an antibiotic from a mold). Atractyloside binds to the translocase when its nucleotide site faces the cytoplasm, whereas bongkrekic acid binds when this site faces the mitochondrial matrix. Oxidative phosphorylation stops soon after either inhibitor is added, showing that ATP-ADP translocase is essential for maintaining adequate amounts of ADP to accept the energy associated with the proton-motive force.

Mitochondrial diseases are being discovered

The number of diseases that can be attributed to mitochondrial mutations is steadily growing in step with our growing understanding of the biochemistry and genetics of mitochondria. The prevalence of mitochondrial diseases is estimated to be from 10 to 15 per 100,000 people, roughly equivalent to the prevalence of the muscular dystrophies. The first mitochondrial disease to be understood was Leber hereditary optic neuropathy (LHON), a form of blindness that strikes in midlife as a result of mutations in Complex I. Some of these mutations impair NADH utilization, whereas others block electron transfer to Q. Mutations in Complex I are the most frequent cause of mitochondrial diseases. The accumulation of mutations in mitochondrial genes in a span of several decades may contribute to aging, degenerative disorders, and cancer.

A human egg harbors several hundred thousand molecules of mitochondrial DNA, whereas a sperm contributes only a few hundred and thus has little effect on the mitochondrial genotype. Because the maternally inherited mitochondria are present in large numbers and not all of the mitochondria may be affected, the pathologies of mitochondrial mutants can be quite complex. Even within a single family carrying an identical mutation, chance fluctuations in the percentage of mitochondria with the mutation lead to large variations in the nature and severity of the symptoms of the pathological condition as well as the time of onset. As the percentage of defective mitochondria increases, energy-generating capacity diminishes until, at some threshold, the cell can no longer function properly. Defects in cellular respiration are doubly dangerous. Not only does energy transduction decrease, but also the likelihood that reactive oxygen species will be generated increases. Organs that are highly dependent on oxidative phosphorylation, such as the nervous system, the retina, and the heart, are most vulnerable to mutations in mitochondrial DNA.

Mitochondria play a key role in apoptosis

In the course of development or in cases of significant cell damage, individual cells within multicellular organisms undergo *programmed cell death,* or *apoptosis.* Mitochondria act as control centers regulating this process. Although the details have not yet been established, the outer membrane of damaged mitochondria becomes highly permeable, a process referred to as *mitochondrial outer membrane permeabilization* (MOMP). This permeabilization is instigated by a family of proteins (Bcl family) that were initially discovered because of their role in cancer. One of the most potent activators of apoptosis, cytochrome *c*, exits the mitochondria and interacts with apoptotic peptidase-activating factor 1 (APAF-1), which leads to the formation of the *apoptosome*. The apoptosome recruits and activates a proteolytic enzyme called *caspase 9*, a member of the cysteine protease family (Section 9.1) that in turn activates a cascade of other caspases. Each caspase type destroys a particular target, such as the proteins that maintain cell structure. Another target is a protein that inhibits an enzyme that destroys DNA (an enzyme called caspase-activated DNAse or CAD), freeing CAD to cleave the genetic material. This cascade of proteolytic enzymes has been called "death by a thousand tiny cuts."

Power transmission by proton gradients is a central motif of bioenergetics

The main concept presented in this chapter is that mitochondrial electron transfer and ATP synthesis are linked by a transmembrane proton gradient. ATP synthesis in bacteria and chloroplasts also is driven by proton gradients. In fact, proton gradients power a variety of energy-requiring processes such as the active transport of calcium ions by mitochondria, the entry of some amino acids and sugars into bacteria, the rotation of bacterial flagella, and the transfer of electrons from $NADP^+$ to NADPH. Proton gradients can also be used to generate heat, as in nonshivering thermogenesis. It is evident that *proton gradients are a central interconvertible currency of free energy in biological systems* (Figure 18.45). Mitchell noted that the proton-motive force is a marvelously simple and effective store of free energy because it requires only a thin, closed lipid membrane between two aqueous phases.

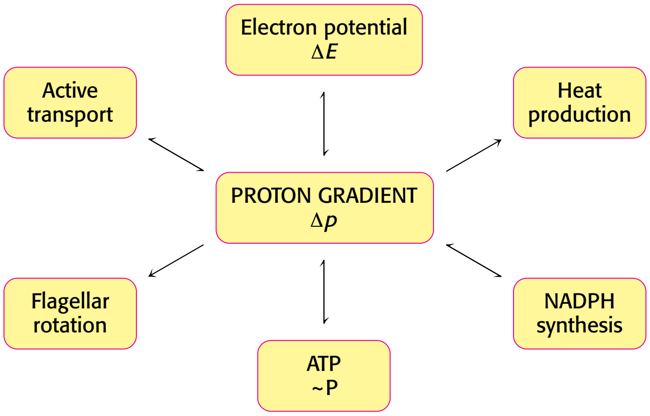


FIGURE 18.45 The proton gradient is an interconvertible form of free energy.

SUMMARY

18.1 Eukaryotic Oxidative Phosphorylation Takes Place in Mitochondria

Mitochondria generate most of the ATP required by aerobic cells through a joint endeavor of the reactions of the citric acid cycle, which take place in the mitochondrial matrix, and oxidative phosphorylation, which takes place in the inner mitochondrial membrane. Mitochondria are descendants of a free-living bacterium that established a symbiotic relation with another cell.

18.2 Oxidative Phosphorylation Depends on Electron Transfer

In oxidative phosphorylation, the synthesis of ATP is coupled to the flow of electrons from NADH or $FADH_2$ to O_2 by a proton gradient across the inner mitochondrial membrane. Electron flow through three asymmetrically oriented transmembrane complexes results in the pumping of protons out of the mitochondrial matrix and the generation of a membrane potential. ATP is synthesized when protons flow back to the matrix through a channel in an ATP-synthesizing complex, called ATP synthase (also known as F_0F_1-ATPase). Oxidative phosphorylation exemplifies a fundamental theme of bioenergetics: the transmission of free energy by proton gradients.

18.3 The Respiratory Chain Consists of Four Complexes: Three Proton Pumps and a Physical Link to the Citric Acid Cycle

The electron carriers in the respiratory assembly of the inner mitochondrial membrane are quinones, flavins, iron–sulfur complexes, heme groups of cytochromes, and copper ions. Electrons from NADH are transferred to the FMN prosthetic group of NADH-Q oxidoreductase (Complex I), the first of four complexes. This oxidoreductase also contains Fe-S centers. The electrons emerge in QH_2, the reduced form of ubiquinone (Q). The citric acid cycle enzyme succinate dehydrogenase is a component of the succinate-Q reductase complex (Complex II), which donates electrons from $FADH_2$ to Q to form QH_2. This hydrophobic carrier transfers its electrons to Q-cytochrome *c* oxidoreductase (Complex III), a complex that contains cytochromes *b* and c_1 and an Fe-S center. This complex reduces cytochrome *c,* a water-soluble peripheral membrane protein. Cytochrome *c* transfers electrons to cytochrome *c* oxidase (Complex IV). This complex contains cytochromes *a* and a_3 and three copper ions. A heme iron ion and a copper ion in this oxidase transfer electrons to O_2, the ultimate acceptor, to form H_2O.

18.4 A Proton Gradient Powers the Synthesis of ATP

The flow of electrons through Complexes I, III, and IV leads to the transfer of protons from the matrix side to the cytoplasmic side of the inner mitochondrial membrane. A proton-motive force consisting of a pH gradient (matrix side basic) and a membrane potential (matrix side negative) is generated. The flow of protons back to the matrix side through ATP synthase drives ATP synthesis. The enzyme complex is a molecular motor made of two operational units: a rotating component and a stationary component. The rotation of the γ subunit induces structural changes in the β subunit that result in the synthesis and release of ATP from the enzyme. Proton influx provides the force for the rotation.

The flow of two electrons through NADH-Q oxidoreductase, Q-cytochrome *c* oxidoreductase, and cytochrome *c* oxidase generates a gradient sufficient to synthesize 1, 0.5, and 1 molecule of ATP, respectively. Hence, 2.5 molecules of ATP are formed per molecule of NADH oxidized in the mitochondrial matrix, whereas only 1.5 molecules of ATP are made per molecule of $FADH_2$ oxidized, because its electrons enter the chain at QH_2, after the first proton-pumping site.

18.5 Many Shuttles Allow Movement Across Mitochondrial Membranes

Mitochondria employ a host of transporters, or carriers, to move molecules across the inner mitochondrial membrane. The electrons of cytoplasmic NADH are transferred into mitochondria by the glycerol phosphate shuttle to form $FADH_2$ from FAD or by the malate–aspartate shuttle to form mitochondrial NADH. The entry of ADP into the mitochondrial matrix is coupled to the exit of ATP by ATP-ADP translocase, a transporter driven by membrane potential.

18.6 The Regulation of Cellular Respiration Is Governed Primarily by the Need for ATP

About 30 molecules of ATP are generated when a molecule of glucose is completely oxidized to CO_2 and H_2O. Electron transport is normally tightly coupled to phosphorylation. NADH and $FADH_2$ are oxidized only if ADP is simultaneously phosphorylated to ATP, a form of regulation called acceptor or respiratory control. Proteins have been identified that uncouple electron transport and ATP synthesis for the generation of heat.

KEY TERMS

oxidative phosphorylation (p. 523)
electron-transport chain (p. 523)
proton-motive force (p. 523)
cellular respiration (p. 524)
reduction (redox, oxidation–reduction, E_0') potential (p. 526)
coenzyme Q (Q, ubiquinone) (p. 530)
Q pool (p. 531)
iron–sulfur (nonheme iron) protein (p. 531)
NADH-Q oxidoreductase (Complex I) (p. 532)
flavin mononucleotide (FMN) (p. 532)
succinate-Q reductase (Complex II) (p. 533)
cytochrome *c* (Cyt *c*) (p. 533)
Q-cytochrome *c* oxidoreductase (Complex III) (p. 534)
Rieske center (p. 534)
Q cycle (p. 535)
cytochrome *c* oxidase (Complex IV) (p. 535)
superoxide dismutase (p. 539)
catalase (p. 540)
ATP synthase (Complex V, F_1F_0 ATPase) (p. 542)
glycerol 3-phosphate shuttle (p. 549)
malate–aspartate shuttle (p. 550)
ATP-ADP translocase (adenine nucleotide translocase, ANT) (p. 550)
respiratory (acceptor) control (p. 553)
uncoupling protein (UCP) (p. 555)
programmed cell death (apoptosis) (p. 557)
mitochondrial outer membrane permeabilization (MOMP) (p. 558)
apoptosome (p. 558)
caspase (p. 558)

PROBLEMS

1. *Breathe or ferment?* Compare fermentation and respiration with respect to electron donors and electron acceptors.

2. *Reference states.* The standard oxidation–reduction potential for the reduction of O_2 to H_2O is given as 0.82 V in Table 18.1. However, the value given in textbooks of chemistry is 1.23 V. Account for this difference.

3. *Less energetic electrons.* Why are electrons carried by $FADH_2$ not as energy rich as those carried by NADH? What is the consequence of this difference?

4. *Now prove it.* Calculate the energy released by the reduction of O_2 with $FADH_2$.

5. *Thermodynamic constraint.* Compare the $\Delta G°'$ values for the oxidation of succinate by NAD^+ and by FAD. Use the data given in Table 18.1 to find the E_0' of the NAD^+-NADH and fumarate-succinate couples, and assume that E_0' for the FAD–$FADH_2$ redox couple is nearly 0.05 V. Why is FAD rather than NAD^+ the electron acceptor in the reaction catalyzed by succinate dehydrogenase?

6. *Give and accept.* Distinguish between an oxidizing agent and a reducing agent.

7. *The benefactor and beneficiary.* Identify the oxidant and the reductant in the following reaction.

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{lactate} + \text{NAD}^+$$

8. *Six of one, half dozen of the other.* How is the redox potential ($\Delta E_0'$) related to the free-energy change of a reaction ($\Delta G°'$)?

9. *Location, location, location.* Iron is a component of many of the electron carriers of the electron-transport chain. How can it participate in a series of coupled redox reactions if the E_0' value is +0.77 V, as seen in Table 18.1?

10. *Line up.* Place the following components of the electron-transport chain in their proper order:

(a) cytochrome *c*
(b) Q-cytochrome *c* oxidoreductase
(c) NADH-Q reductase
(d) cytochrome *c* oxidase
(e) ubiquinone

11. *Like mac and cheese.* Match each term with its description.

(a) Respiration _____
(b) Redox potential _____
(c) Electron-transport chain _____
(d) Flavin mononucleotide (FMN) _____
(e) Iron–sulfur protein _____
(f) Coenzyme Q _____
(g) Cytochrome *c* _____
(h) Q cycle _____
(i) Superoxide dismutase _____
(j) Catalase _____

1. Converts reactive oxygen species into hydrogen peroxide
2. Electron flow from NADH and $FADH_2$ to O_2
3. Facilitates electron flow from FMN to coenzyme Q in Complex I
4. An ATP-generating process in which an inorganic compound serves as the final electron acceptor
5. Measure of the tendency to accept or donate electrons
6. Converts hydrogen peroxide into oxygen and water
7. Funnels electrons from a two-electron carrier to a one-electron carrier
8. Lipid-soluble electron carrier
9. Donates electrons to Complex IV
10. Accepts electrons from NADH in Complex I

12. *Match 'em.*

(a) Complex I ___ 1. Q-cytochrome c oxidoreductase
(b) Complex II ___ 2. Coenzyme Q
(c) Complex III ___ 3. Succinate-Q reductase
(d) Complex IV ___ 4. NADH-Q oxidoreductase
(e) Ubiquinone ___ 5. Cytochrome *c* oxidase

13. *Structural considerations.* Explain why coenzyme Q is an effective mobile electron carrier in the electron-transport chain.

14. *Inhibitors.* Rotenone inhibits electron flow through NADH-Q oxidoreductase. Antimycin A blocks electron flow between cytochromes *b* and c_1. Cyanide blocks electron flow through cytochrome oxidase to O_2. Predict the relative oxidation–reduction state of each of the following respiratory-chain components in mitochondria that are treated with each of the inhibitors:

(a) NAD^+
(b) NADH-Q oxidoreductase
(c) coenzyme Q
(d) cytochrome c_1
(e) cytochrome *c*
(f) cytochrome *a*

15. *Rumored to be a favorite of Elvis.* Amytal is a barbiturate sedative that inhibits electron flow through Complex I. How would the addition of amytal to actively respiring mitochondria affect the relative oxidation–reduction states of the components of the electron-transport chain and the citric acid cycle?

16. *Efficiency.* What is the advantage of having Complexes I, III, and IV associated with one another in the form of a respirasome?

17. *Linked In.* What citric acid cycle enzyme is also a component of the electron-transport chain?

18. *ROS, not ROUS.* What are the reactive oxygen species and why are they especially dangerous to cells?

19. *Reclaim resources.* Humans have only about 250 g of ATP, but even a couch potato needs about 83 kg of ATP to open the bag of chips and use the remote. How is this discrepancy between requirements and resources reconciled?

20. *Energy harvest.* What is the yield of ATP when each of the following substrates is completely oxidized to CO_2 by a mammalian cell homogenate? Assume that glycolysis, the citric acid cycle, and oxidative phosphorylation are fully active.

(a) Pyruvate
(b) Lactate
(c) Fructose 1,6-bisphosphate
(d) Phosphoenolpyruvate
(e) Galactose
(f) Dihydroxyacetone phosphate

21. *Potent poisons.* What is the effect of each of the following inhibitors on electron transport and ATP formation by the respiratory chain?

(a) Azide
(b) Atractyloside
(c) Rotenone
(d) DNP
(e) Carbon monoxide
(f) Antimycin A

22. *A question of coupling.* What is the mechanistic basis for the observation that the inhibitors of ATP synthase also lead to an inhibition of the electron-transport chain?

23. *A Brownian ratchet wrench.* What causes the **c** subunits of ATP synthase to rotate? What determines the direction of rotation?

24. *An essential residue.* The conduction of protons by the F_0 unit of ATP synthase is blocked by dicyclohexylcarbodiimide, which reacts readily with carboxyl groups. What are the most likely targets of action of this reagent? How might you use site-specific mutagenesis to determine whether this residue is essential for proton conduction?

25. *Alternative routes.* The most common metabolic sign of mitochondrial disorders is lactic acidosis. Why?

26. *Connections.* How does the inhibition of ATP-ADP translocase affect the citric acid cycle? Glycolysis?

27. *O_2 consumption.* Oxidative phosphorylation in mitochondria is often monitored by measuring oxygen consumption. When oxidative phosphorylation is proceeding rapidly, the mitochondria will rapidly consume oxygen. If there is little oxidative phosphorylation, only small amounts of oxygen will be used. You are given a suspension of isolated mitochondria and directed to add the following compounds in the order from *a* to *h*. With the addition of each compound, all of the previously added compounds remain present. Predict the effect of each addition on oxygen consumption by the isolated mitochondria.

(a) Glucose
(b) ADP + P_i
(c) Citrate
(d) Oligomycin
(e) Succinate
(f) Dinitrophenol
(g) Rotenone
(h) Cyanide

28. *P : O ratios.* The number of molecules of inorganic phosphate incorporated into organic form per atom of oxygen consumed, termed the *P : O ratio,* was frequently used as an index of oxidative phosphorylation.

(a) What is the relation of the P : O ratio to the ratio of the number of protons translocated per electron pair ($H^+/2\ e^-$) and the ratio of the number of protons needed to synthesize ATP and transport it to the cytoplasm (P/H^+)?

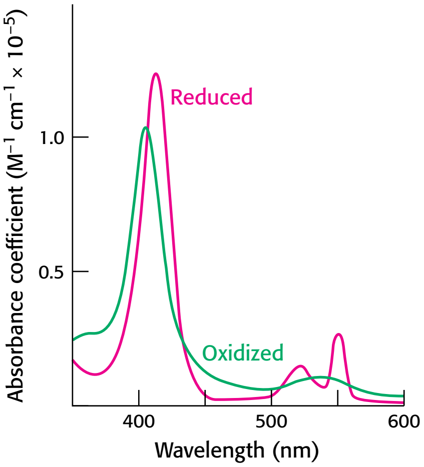
(b) What are the P : O ratios for electrons donated by matrix NADH and by succinate?

29. *Cyanide antidote.* The immediate administration of nitrite is a highly effective treatment for cyanide poisoning. What is the basis for the action of this antidote? (Hint: Nitrite oxidizes ferrohemoglobin to ferrihemoglobin.)

30. *Runaway mitochondria 1.* Suppose that the mitochondria of a patient oxidize NADH irrespective of whether ADP is present. The P : O ratio for oxidative phosphorylation by these mitochondria is less than normal. Predict the likely symptoms of this disorder.

31. *Recycling device.* The cytochrome *b* component of Q-cytochrome *c* oxidoreductase enables both electrons of QH_2 to be effectively utilized in generating a proton-motive force. Cite another recycling device in metabolism that brings a potentially dead end reaction product back into the mainstream.

32. *Crossover point.* The precise site of action of a respiratory-chain inhibitor can be revealed by the *crossover technique*. Britton Chance devised elegant spectroscopic methods for determining the proportions of the oxidized and reduced forms of each carrier. This determination is feasible because the forms have distinctive absorption spectra, as illustrated in the adjoining graph for cytochrome *c*. You are given a new inhibitor and find that its addition to respiring mitochondria causes the carriers between NADH and QH_2 to become more reduced and those between cytochrome *c* and O_2 to become more oxidized. Where does your inhibitor act?



33. *Runaway mitochondria 2.* Years ago, uncouplers were suggested to make wonderful diet drugs. Explain why this idea was proposed and why it was rejected. Why might the producers of antiperspirants be supportive of the idea?

34. *Everything is connected.* If actively respiring mitochondria are exposed to an inhibitor of ATP-ADP translocase, the electron-transport chain ceases to operate. Why?

35. *Identifying the inhibition.* You are asked to determine whether a chemical is an electron-transport-chain inhibitor or an inhibitor of ATP synthase. Design an experiment to make this determination.

36. *To each according to its needs.* It has been noted that the mitochondria of muscle cells often have more cristae than the mitochondria of liver cells. Provide an explanation for this observation.

37. *Opposites attract.* An arginine residue (Arg 210) in the **a** subunit of the *E. coli* ATP synthase is near the aspartate residue (Asp 61) in the matrix-side proton channel. How might Arg 210 assist proton flow?

38. *Variable c subunits.* Recall that the number of **c** subunits in the **c** ring appears to range between 8 and 14. This number is significant because it determines the number of protons that must be transported to generate a molecule of ATP. Each 360-degree rotation of the γ subunit leads to the synthesis and release of three molecules of ATP. Thus, if there are 10 **c** subunits in the ring (as was observed in a crystal structure of yeast mitochondrial ATP synthase), each ATP generated requires the transport of $10/3 = 3.33$ protons. How many protons are required to form ATP if the ring has 12 **c** subunits? 14?

39. *Counterintuitive.* Under some conditions, mitochondrial ATP synthase has been observed to actually run in reverse. How would that situation affect the proton-motive force?

40. *Etiology? What does that mean?* What does the fact that rotenone appears to increase the susceptibility to Parkinson disease indicate about the etiology of Parkinson disease?

41. *Exaggerating the difference.* Why must ATP-ADP translocase (also called adenine nucleotide translocase or ANT) use Mg^{2+}-free forms of ATP and ADP?

42. *Respiratory control.* The rate of oxygen consumption by mitochondria increases markedly when ADP is added and then returns to its initial value when the added ADP has been converted into ATP (Figure 18.39). Why does the rate decrease?

43. *Same, but different.* Why is the electroneutral exchange of $H_2PO_4^-$ for OH^- indistinguishable from the electroneutral symport of $H_2PO_4^-$ and H^+?

44. *Multiple uses.* Give an example of the use of the proton-motive force in ways other than for the synthesis of ATP.

Chapter Integration Problems

45. *Just obeying the laws.* Why do isolated F_1 subunits of ATP synthase catalyze ATP hydrolysis?

46. *The right location.* Some cytoplasmic kinases, enzymes that phosphorylate substrates at the expense of ATP, bind to voltage-dependent anion channels. What might the advantage of this binding be?

47. *No exchange.* Mice that completely lack ATP-ADP translocase (ANT^-/ANT^-) can be made by using the knockout technique. Remarkably, these mice are viable but have the following pathological conditions: (1) high serum levels of

lactate, alanine, and succinate; (2) little electron transport; and (3) a six- to eightfold increase in the level of mitochondrial H_2O_2 compared with that in normal mice. Provide a possible biochemical explanation for each of these conditions.

48. *Maybe you shouldn't take your vitamins.* Exercise is known to increase insulin sensitivity and to ameliorate type 2 diabetes (Chapter 27). Recent research suggests that taking antioxidant vitamins might mitigate the beneficial effects of exercise with respect to ROS protection.

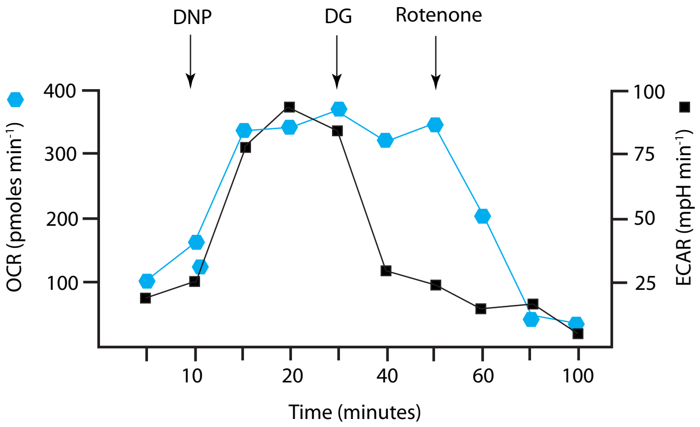
(a) What are the antioxidant vitamins?

(b) How does exercise protect against ROS?

(c) Explain why vitamins might counteract the effects of exercise.

Chapter Integration and Data Interpretation Problem

49. *Monitoring energy sources.* XF technology (Seahorse Bioscience) now allows the measurement of the rate of aerobic respiration and lactic acid fermentation simultaneously in real time in cultured cells. The extent of aerobic respiration is determined by measuring the oxygen consumption rate (OCR, measured in picomoles of oxygen consumed per minute) while the rate of glycolysis correlates with the extracellular acidification rate [ECAR-milli pH per minute (the changes in pH that occur over time)]. The graph below shows the results of an experiment using the new technology.



Dinitrophenol (DNP), the glycolysis inhibitor 2-deoxyglucose (DG), and rotenone were added sequentially to cell cultures.

(a) What is the effect on OCR and ECAR of adding DNP to the cell culture? Explain these results.

(b) Explain the effect of the addition of 2-deoxyglucose.

(c) Explain how 2-deoxyglucose acts as an inhibitor of glycolysis?

(d) Explain the effect of the addition of rotenone.

Data Interpretation Problem

50. *Mitochondrial disease.* A mutation in a mitochondrial gene encoding a component of ATP synthase has been identified. People who have this mutation suffer from muscle weakness, ataxia, and retinitis pigmentosa. A tissue biopsy was performed on each of three patients having this mutation, and submitochondrial particles were isolated that were capable of succinate-sustained ATP synthesis. First, the activity of the ATP synthase was measured on the addition of succinate and the following results were obtained.

	ATP synthase activity (nmol of ATP formed min^{-1} mg^{-1})
Controls	3.0
Patient 1	0.25
Patient 2	0.11
Patient 3	0.17

(a) What was the purpose of the addition of succinate?

(b) What is the effect of the mutation on succinate-coupled ATP synthesis?

Next, the ATPase activity of the enzyme was measured by incubating the submitochondrial particles with ATP in the absence of succinate.

	ATP hydrolysis (nmol of ATP hydrolyzed min^{-1} mg^{-1})
Controls	33
Patient 1	30
Patient 2	25
Patient 3	31

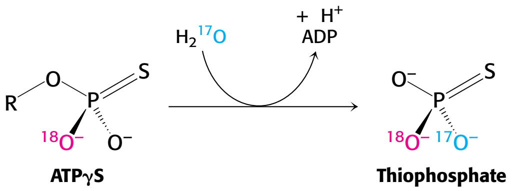
(c) Why was succinate omitted from the reaction?

(d) What is the effect of the mutation on ATP hydrolysis?

(e) What do these results, in conjunction with those obtained in the first experiment, tell you about the nature of the mutation?

Mechanism Problem

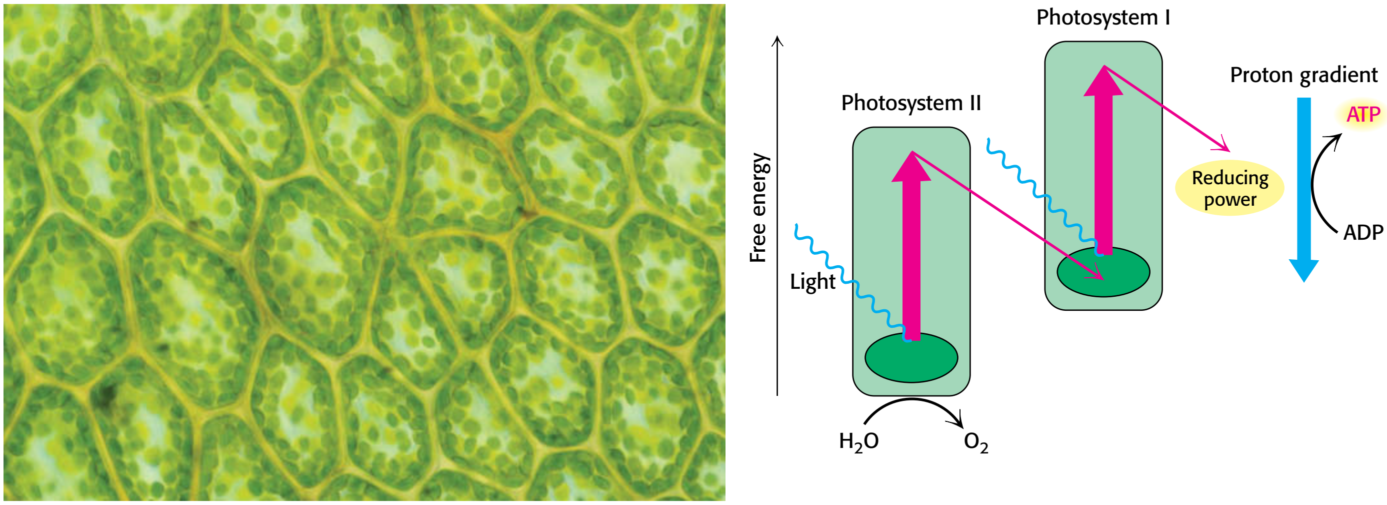
51. *Chiral clue.* ATPγS, a slowly hydrolyzed analog of ATP, can be used to probe the mechanism of phosphoryl-transfer reactions. Chiral ATPγS has been synthesized containing ^{18}O in a specific γ position and ordinary ^{16}O elsewhere in the molecule. The hydrolysis of this chiral molecule by ATP synthase in ^{17}O-enriched water yields inorganic [^{16}O,^{17}O,^{18}O]thiophosphate having the following absolute configuration. In contrast, the hydrolysis of this chiral ATPγS by a calcium-pumping ATPase from muscle gives thiophosphate of the opposite configuration. What is the simplest interpretation of these data?



CHAPTER

19

The Light Reactions of Photosynthesis



Chloroplasts (left) convert light energy into chemical energy. High-energy electrons in chloroplasts are transported through two photosystems (right). In this transit, which culminates in the generation of reducing power, ATP is synthesized in a manner analogous to mitochondrial ATP synthesis. In contrast with mitochondrial electron transport, however, electrons in chloroplasts are energized by light. [(Left) John Durham/ Science Source.]

Every year, Earth is bathed in photons with a total energy content of approximately 10^{24} kJ. By comparison, a major hurricane has 10^{-9} that amount of energy. The source of the energy is, of course, the electromagnetic radiation from the sun. On our planet, there are organisms capable of collecting a fraction, approximately 1%, of this solar energy and converting it into chemical energy. Green plants are the most obvious of these organisms, although 60% of this conversion is carried out by algae and bacteria. Although only a small amount of the incident energy is captured, it is enough to power all life on Earth. This transformation is perhaps the most important of all of the energy transformations that we will see in our study of biochemistry; without it, life as we know it on our planet simply could not exist.

The process of converting electromagnetic radiation into chemical energy is called *photosynthesis,* which uses light energy to convert carbon dioxide and water into carbohydrates and oxygen.

$$CO_2 + H_2O \xrightarrow{\text{Light}} (CH_2O) + O_2$$

In this equation, CH_2O represents carbohydrate, primarily sucrose and starch. Photosynthetic carbohydrate synthesis is the most common metabolic pathway on Earth. These carbohydrates provide not only the energy to

OUTLINE

19.1 Photosynthesis Takes Place in Chloroplasts
19.2 Light Absorption by Chlorophyll Induces Electron Transfer
19.3 Two Photosystems Generate a Proton Gradient and NADPH in Oxygenic Photosynthesis
19.4 A Proton Gradient across the Thylakoid Membrane Drives ATP Synthesis
19.5 Accessory Pigments Funnel Energy into Reaction Centers
19.6 The Ability to Convert Light into Chemical Energy Is Ancient

run the biological world, but also the carbon molecules to make a wide array of biomolecules. Photosynthetic organisms are called *autotrophs* (literally, "self-feeders") because they can synthesize chemical fuels such as glucose from carbon dioxide and water by using sunlight as an energy source and then recover some of this energy from the synthesized glucose through the glycolytic pathway and aerobic metabolism. Organisms that obtain energy from chemical fuels only are called *heterotrophs,* which ultimately depend on autotrophs for their fuel.

Photosynthesis is composed of two parts: the light reactions and the dark reactions. In the *light reactions,* light energy is transformed into two forms of biochemical energy with which we are already familiar: reducing power and ATP. The products of the light reactions are then used in the dark reactions to drive the reduction of CO_2 and its conversion into glucose and other sugars. The dark reactions are also called the *Calvin cycle* or *light-independent reactions* and will be discussed in Chapter 20.

Photosynthesis converts light energy into chemical energy

The light reactions of photosynthesis closely resemble the events of oxidative phosphorylation. In Chapters 17 and 18, we learned that cellular respiration is the oxidation of glucose to CO_2 with the reduction of O_2 to water, a process that generates ATP. In photosynthesis, this process must be reversed—reducing CO_2 and oxidizing H_2O to synthesize glucose.

$$\text{Energy} + 6\,H_2O + 6\,CO_2 \xrightarrow{\text{Photosynthesis}} C_6H_{12}O_6 + 6\,O_2$$

$$C_6H_{12}O_6 + 6\,O_2 \xrightarrow[]{\substack{\text{Cellular}\\ \text{respiration}}} 6\,O_2 + 6\,H_2O + \text{energy}$$

Although the processes of respiration and photosynthesis are chemically opposite each other, the biochemical principles governing the two processes are nearly identical. The key to both processes is the generation of high-energy electrons. The citric acid cycle oxidizes carbon fuels to CO_2 to generate high-energy electrons. The flow of these high-energy electrons down an electron-transport chain generates a proton-motive force. This proton-motive force is then transduced by ATP synthase to form ATP. To synthesize glucose from CO_2, high-energy electrons are required for two purposes: (1) to provide reducing power in the form of NADPH to reduce CO_2, and (2) to generate ATP to power the reduction. How can high-energy electrons be generated without using a chemical fuel? *Photosynthesis uses energy from light to boost electrons from a low-energy state to a high-energy state*. In the high-energy, unstable state, nearby molecules can abscond with the excited electrons. These electrons are used to produce reducing power, and to generate a proton-motive force across a membrane, which subsequently drives the synthesis of ATP. The reactions that are powered by sunlight are called the *light reactions* (Figure 19.1).

Photosynthetic yield

"If a year's yield of photosynthesis were amassed in the form of sugar cane, it would form a heap over two miles high and with a base 43 square miles."

—G. E. Fogge

If all of this sugar cane were converted into sugar cubes (0.5 inch, or 1.27 cm, on a side) and stacked end to end, the sugar cubes would extend 1.6×10^{10} miles (2.6×10^{10} kilometers) or to the dwarf planet Pluto.

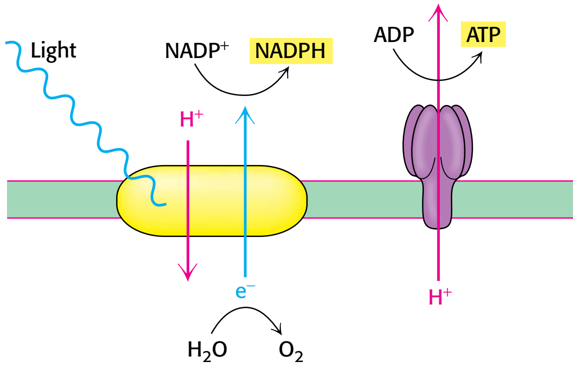


FIGURE 19.1 The light reactions of photosynthesis. Light is absorbed and the energy is used to drive electrons from water to generate NADPH and to drive protons across a membrane. These protons return through ATP synthase (purple structure) to make ATP.

Photosynthesis in green plants is mediated by two kinds of light reactions. Photosystem I generates reducing power in the form of NADPH, but, in the process, becomes electron deficient. Photosystem II oxidizes water and transfers the electrons to replenish the electrons lost by photosystem I. A side product of these reactions is O_2. Electron flow from photosystem II to photosystem I generates the transmembrane proton gradient, augmented by the protons released by the oxidation of water, that drives the synthesis of ATP. In keeping with the similarity of their principles of operation, both processes take place in double-membrane organelles: mitochondria for cellular respiration and chloroplasts for photosynthesis.

Photosynthetic catastrophe
If photosynthesis were to cease, all higher forms of life would be extinct in about 25 years. A milder version of such a catastrophe ended the Cretaceous period 65.1 million years ago when a large asteroid struck the Yucatan Peninsula of Mexico. Enough dust was sent into the atmosphere that photosynthetic capacity was greatly diminished, which apparently led to the disappearance of the dinosaurs and allowed the mammals to rise to prominence.

19.1 Photosynthesis Takes Place in Chloroplasts

Photosynthesis, the means of converting light into chemical energy, takes place in organelles called *chloroplasts,* typically 5 μm long. Like a mitochondrion, a chloroplast has an outer membrane and an inner membrane, with an intervening intermembrane space (Figure 19.2). The inner membrane surrounds a space called the *stroma,* which is the site of the dark reactions of photosynthesis (Section 20.1). In the stroma are membranous structures called *thylakoids,* which are flattened sacs, or discs. The thylakoid sacs are stacked to form a *granum*. Different grana are linked by regions of thylakoid membrane called *stroma lamellae* (Figure 19.3). The thylakoid membranes separate the thylakoid space from the stroma space. Thus, chloroplasts have three different membranes (*outer, inner*, and *thylakoid membranes*) and three separate spaces (*intermembrane, stroma,* and *thylakoid spaces*). In developing chloroplasts, thylakoids arise from budding of the inner membrane, and so they are analogous to the mitochondrial cristae. Like the mitochondrial cristae, they are the site of coupled oxidation–reduction reactions of the light reactions that generate the proton-motive force.

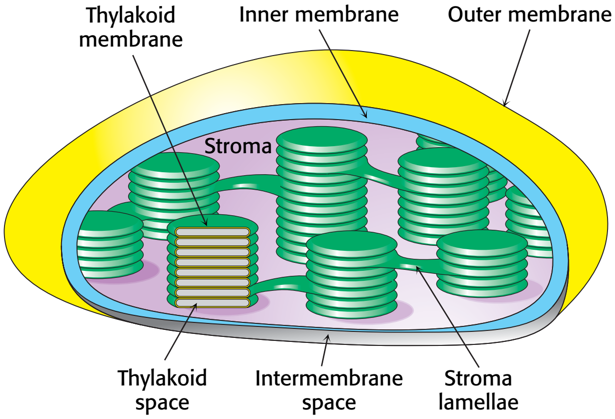


FIGURE 19.2 Diagram of a chloroplast.

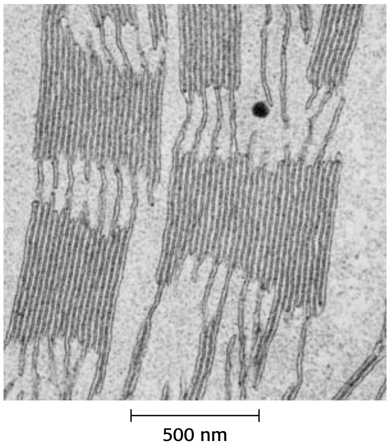


FIGURE 19.3 Electron micrograph of a chloroplast from a spinach leaf. The thylakoid membranes pack together to form grana. [Courtesy of Dr. Kenneth Miller.]

The primary events of photosynthesis take place in thylakoid membranes

The thylakoid membranes contain the energy-transforming machinery: light-harvesting proteins, reaction centers, electron-transport chains, and ATP synthase. These membranes contain nearly equal amounts of lipids and proteins. The lipid composition is highly distinctive: about 75% of the total lipids are *galactolipids* and 10% are *sulfolipids*, whereas only 10% are phospholipids. The thylakoid membrane and the inner membrane, like the inner mitochondrial membrane, are impermeable to most molecules and ions. The outer membrane of a chloroplast, like that of a mitochondrion, is highly permeable to small molecules and ions. The stroma contains the soluble enzymes that utilize the NADPH and ATP synthesized by the thylakoids to convert CO_2 into sugar. Plant-leaf cells contain between 1 and 100 chloroplasts, depending on the species, cell type, and growth conditions.

Chloroplasts arose from an endosymbiotic event

Chloroplasts contain their own DNA and the machinery for replicating and expressing it. However, chloroplasts are not autonomous: they also contain many proteins encoded by nuclear DNA. How did the intriguing relation between the cell and its chloroplasts develop? We now believe that, in a manner analogous to the evolution of mitochondria (Section 18.1), chloroplasts are the result of endosymbiotic events in which a photosynthetic microorganism, most likely an ancestor of a cyanobacterium (Figure 19.4), was engulfed by a eukaryotic host. Evidence suggests that chloroplasts in higher plants and green algae are derived from a single endosymbiotic event, whereas those in red and brown algae are derived from at least one additional event.

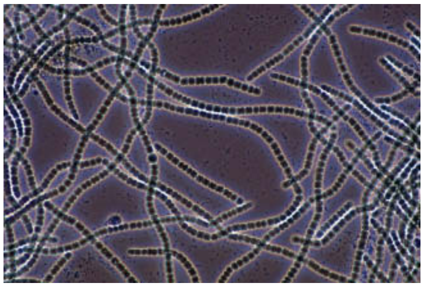


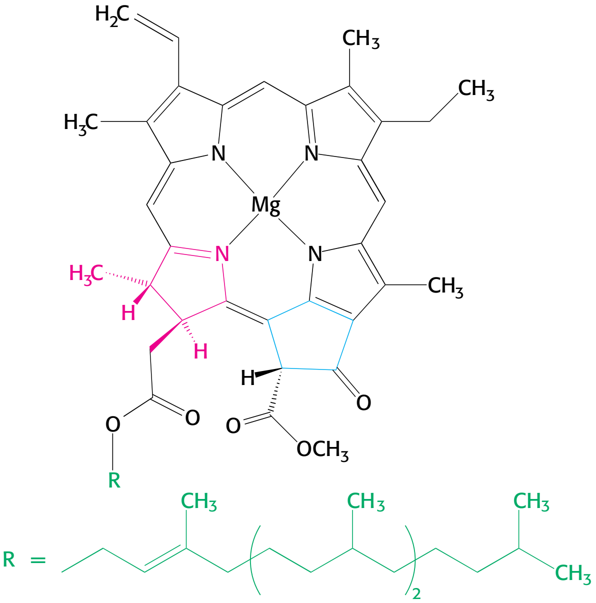
FIGURE 19.4 Cyanobacteria. A colony of the photosynthetic filamentous cyanobacterium *Anabaena* shown at 450× magnification. Ancestors of these bacteria are thought to have evolved into present-day chloroplasts. [Michael Abbey/Science Source.]

The chloroplast genome is smaller than that of a cyanobacterium, but the two genomes have key features in common. Both are circular and have a single start site for DNA replication, and the genes of both are arranged in operons—sequences of functionally related genes under common control (Chapter 31). In the course of evolution, many of the genes of the chloroplast ancestor were transferred to the plant cell's nucleus or, in some cases, lost entirely, thus establishing a fully dependent relation.

19.2 Light Absorption by Chlorophyll Induces Electron Transfer

The trapping of light energy is the key to photosynthesis. The first event is the absorption of light by a photoreceptor molecule. The principal photoreceptor in the chloroplasts of most green plants is the pigment molecule *chlorophyll* a, a substituted tetrapyrrole (Figure 19.5). The four nitrogen atoms of

FIGURE 19.5 Chlorophyll. Like heme, chlorophyll *a* is a cyclic tetrapyrrole. One of the pyrrole rings (shown in red) is reduced, and an additional five-carbon ring (shown in blue) is fused to another pyrrole ring. A phytol chain (shown in green) is connected by an ester linkage. Magnesium ion binds at the center of the structure.



the pyrroles are coordinated to a magnesium ion. Unlike a porphyrin such as heme, chlorophyll has a reduced pyrrole ring and an additional 5-carbon ring fused to one of the pyrrole rings. Another distinctive feature of chlorophyll is the presence of *phytol,* a highly hydrophobic 20-carbon alcohol, esterified to an acid side chain.

Chlorophylls are very effective photoreceptors because they contain networks of conjugated double bonds—alternating single and double bonds. Such compounds are called conjugated *polyenes*. In polyenes, the electrons are not localized to a particular atomic nucleus. Upon the absorption of light energy, the electron transitions from a low energy molecular orbital to a higher energy orbital. Chlorophylls have very strong absorption bands in the visible region of the spectrum, where the solar output reaching Earth is maximal (Figure 19.6). Chlorophyll *a*'s peak molar extinction coefficient (ε), a measure of a compound's ability to absorb light, is greater than $10^5\ \text{M}^{-1}\ \text{cm}^{-1}$, among the highest observed for organic compounds.

What happens when light is absorbed by a pigment molecule? The energy from the light excites an electron from its ground energy level to an excited energy level (Figure 19.7). This high-energy electron can have one of two fates. For most compounds that absorb light, the electron simply returns to the ground state and the absorbed energy is converted into heat. However, if a suitable electron acceptor is nearby, as is the case for chlorophyll in photosynthetic systems, the excited electron can move from the initial molecule to the acceptor (Figure 19.8). A positive charge forms on the initial molecule, owing to the loss of an electron, and a negative charge forms on the acceptor, owing to the gain of an electron. Hence, this process is referred to as *photoinduced charge separation*.

In chloroplasts, the site at which the charge separation takes place within each photosystem is called the *reaction center*. The photosynthetic apparatus is arranged to maximize photoinduced charge separation and minimize an unproductive return of the electron to its ground state. The electron, extracted from its initial site by the absorption of light, now has reducing power: it can reduce other molecules to store the energy originally obtained from light in chemical forms.

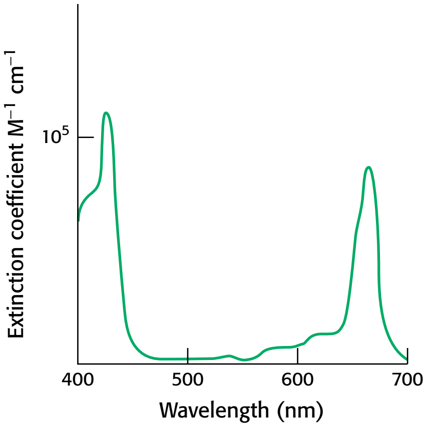


FIGURE 19.6 Light absorption by chlorophyll *a*. Chlorophyll *a* absorbs visible light efficiently, as judged by the extinction coefficient near $10^5\ \text{M}^{-1}\ \text{cm}^{-1}$.

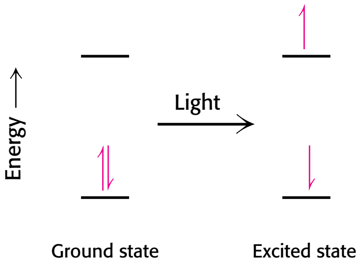


FIGURE 19.7 Light absorption. The absorption of light leads to the excitation of an electron from its ground state to a higher energy level.

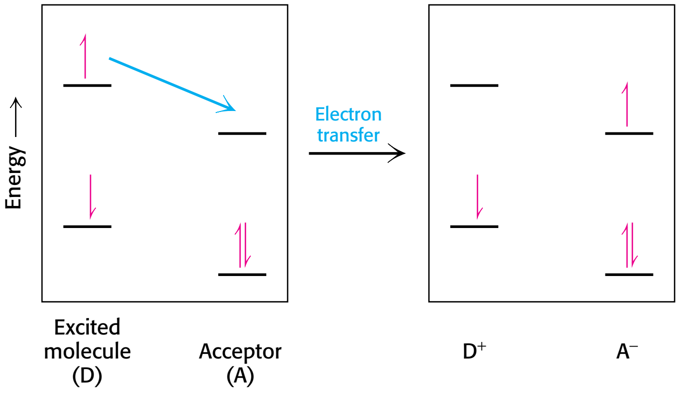


FIGURE 19.8 Photoinduced charge separation. If a suitable electron acceptor is nearby, an electron that has been moved to a high energy level by light absorption can move from the excited molecule to the acceptor.

A special pair of chlorophylls initiate charge separation

Photosynthetic bacteria such as *Rhodopseudomonas viridis* contain a photosynthetic reaction center that has been revealed at atomic resolution.

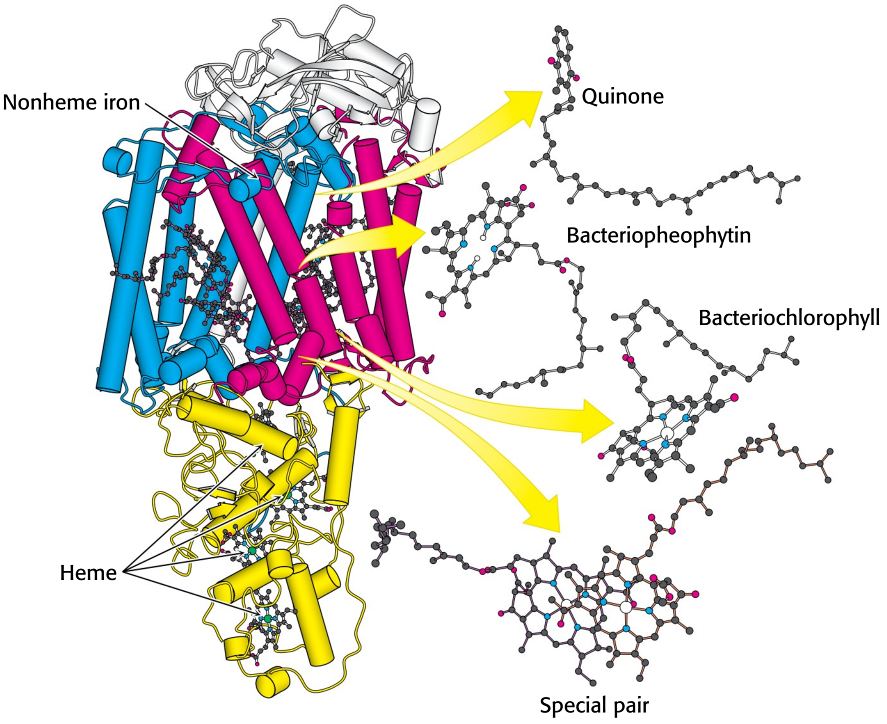
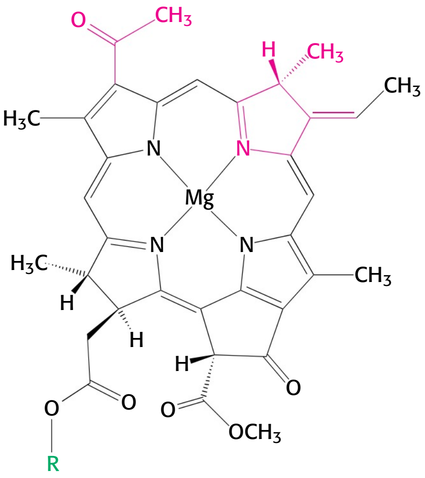
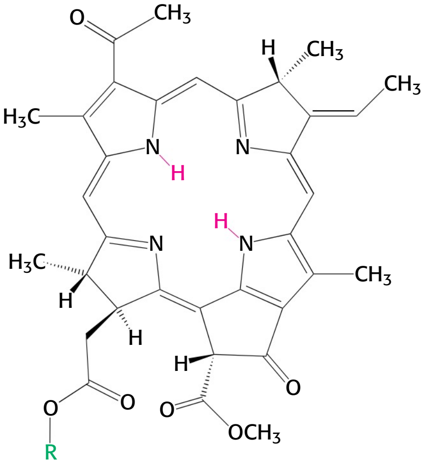


FIGURE 19.9 Bacterial photosynthetic reaction center. The core of the reaction center from *Rhodopseudomonas viridis* consists of two similar chains: L (red) and M (blue). An H chain (white) and a cytochrome subunit (yellow) complete the structure. *Notice* that the L and M subunits are composed largely of α helices that span the membrane. *Also notice* that a chain of electron-carrying prosthetic groups, beginning with a special pair of bacteriochlorophylls and ending at a bound quinone, runs through the structure from bottom to top in this view. [Drawn from 1PRC.pdb.]



Bacteriochlorophyll *b*
(BChl-*b*)



Bacteriopheophytin
(BPh)

The bacterial reaction center consists of four polypeptides: L (31 kDa, red), M (36 kDa, blue), and H (28 kDa, white) subunits and C, a *c*-type cytochrome with four *c*-type hemes (yellow) (Figure 19.9). *Sequence comparisons and low-resolution structural studies have revealed that the bacterial reaction center is homologous to the more complex plant systems*. Thus, many of our observations of the bacterial system apply to plant systems as well.

The L and M subunits form the structural and functional core of the bacterial photosynthetic reaction center (Figure 19.9). Each of these homologous subunits contains five transmembrane helices, in contrast with the H subunit, which has just one. The H subunit lies on the cytoplasmic side of the cell membrane, and the cytochrome subunit lies on the exterior face of the cell membrane, called the *periplasmic side* because it faces the periplasm, the space between the cell membrane and the cell wall. Four bacteriochlorophyll *b* (BChl-*b*) molecules, two bacteriopheophytin *b* (BPh) molecules, two quinones (Q_A and Q_B), and a ferrous ion are associated with the L and M subunits.

Bacteriochlorophylls are photoreceptors similar to chlorophylls, except for the reduction of an additional pyrrole ring and other minor differences that shift their absorption maxima to the near infrared, to wavelengths as long as 1000 nm. *Bacteriopheophytin* is the term for a bacteriochlorophyll that has two protons instead of a magnesium ion at its center.

The reaction begins with light absorption by a pair of BChl-*b* molecules that lie near the periplasmic side of the membrane in the L–M dimer. The pair of BChl-*b* molecules is called the *special pair* because of

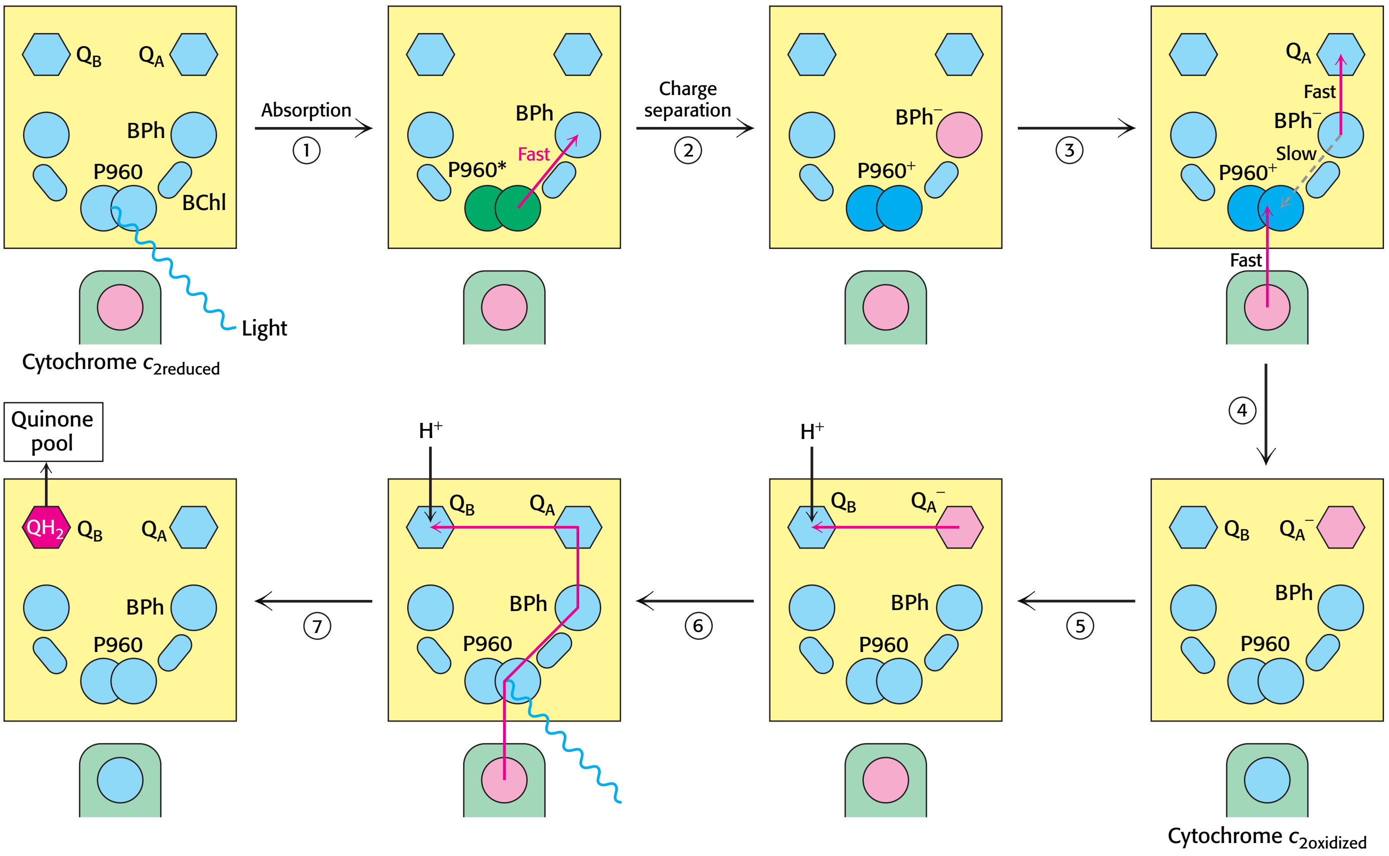


FIGURE 19.10 Electron chain in the photosynthetic bacterial reaction center. The absorption of light by the special pair (P960) results in the rapid transfer of an electron from this site to a bacteriopheophytin (BPh), creating a photoinduced charge separation (steps 1 and 2). (The asterisk on P960 stands for excited state.) The possible return of the electron from the pheophytin to the oxidized special pair is suppressed by the "hole" in the special pair being refilled with an electron from the cytochrome subunit and the electron from the pheophytin being transferred to a quinone (Q_A) that is farther away from the special pair (steps 3 and 4). Q_A passes the electron to Q_B. The reduction of a quinone (Q_B) on the cytoplasmic side of the membrane results in the uptake of two protons from the cytoplasm (steps 5 and 6). The reduced quinone can move into the quinone pool in the membrane (step 7).

its fundamental role in photosynthesis. The special pair absorbs light maximally at 960 nm, and, for this reason, is often called *P960* (*P* stands for pigment). After absorbing light, the excited special pair ejects an electron, which is transferred through another BChl-*b* to a bacteriopheophytin (Figure 19.10, steps 1 and 2). This initial charge separation yields a positive charge on the special pair ($P960^+$) and a negative charge on BPh (BPh^-). The electron ejection and transfer take place in less than 10 picoseconds (10^{-11} s).

A nearby electron acceptor, a tightly bound quinone (Q_A), quickly grabs the electron away from BPh^- before the electron has a chance to fall back to the P960 special pair. From Q_A, the electron moves to a more loosely associated quinone, Q_B. The absorption of a second photon and the movement of a second electron from the special pair through the bacteriopheophytin to the quinones completes the two-electron reduction of Q_B from Q to QH_2. Because the Q_B-binding site lies near the cytoplasmic side of the membrane, *two protons are taken up from the cytoplasm, contributing to the development of a proton gradient across the cell membrane* (Figure 19.10, steps 5, 6, and 7).

In their high-energy states, $P960^+$ and BPh^- could undergo charge recombination; that is, the electron on BPh^- could move back to neutralize the positive charge on the special pair. Its return to the special pair would waste a valuable high-energy electron and simply convert the absorbed light energy into heat. How is charge recombination prevented? Two factors in the structure of the reaction center work together to suppress charge recombination nearly completely (Figure 19.10, steps 3 and 4). First, the next electron acceptor (Q_A) is less than 10 Å away from BPh^-, and so the electron is rapidly transferred farther away from the special pair. Second, one of the hemes of the cytochrome subunit is less than 10 Å away from the special pair, and so the positive charge on P960 is neutralized by the transfer of an electron from the reduced cytochrome.

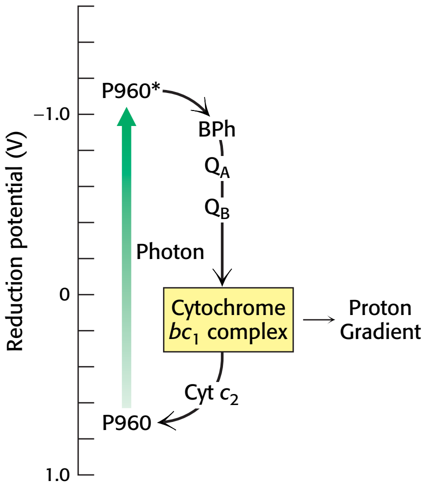


FIGURE 19.11 Cyclic electron flow in the bacterial reaction center. Excited electrons from the P960 reaction center flow through bacteriopheophytin (BPh), a pair of quinone molecules (Q_A and Q_B), cytochrome bc_1 complex, and finally through cytochrome c_2 to the reaction center. The cytochrome bc_1 complex pumps protons as a result of electron flow, which powers the formation of ATP.

Cyclic electron flow reduces the cytochrome of the reaction center

The cytochrome subunit of the reaction center must regain an electron to complete the cycle. It does so by taking back two electrons from reduced quinone (QH_2). QH_2 first enters the Q pool in the membrane where it is reoxidized to Q by complex bc_1, which is homologous to complex III of the respiratory electron-transport chain. Complex bc_1 transfers the electrons from QH_2 to cytochrome c_2, a water-soluble protein in the periplasm, and in the process pumps protons into the periplasmic space. The electrons now on cytochrome c_2 flow to the cytochrome subunit of the reaction center. The flow of electrons is thus cyclic (Figure 19.11). The proton gradient generated in the course of this cycle drives the synthesis of ATP through the action of ATP synthase.

19.3 Two Photosystems Generate a Proton Gradient and NADPH in Oxygenic Photosynthesis

Photosynthesis is more complicated in green plants than in photosynthetic bacteria. In green plants, photosynthesis depends on the interplay of two kinds of membrane-bound, light-sensitive complexes—*photosystem I* (PS I) and *photosystem II* (PS II), as shown in Figure 19.12. There are similarities in photosynthesis between green plants and photosynthetic bacteria. Both require light to energize reaction centers consisting of special pairs, called *P700* for photosystem I and *P680* for photosystem II, and both transfer electrons by using electron-transport chains. However, in plants, electron flow is not cyclic but progresses from photosystem II to photosystem I under most circumstances.

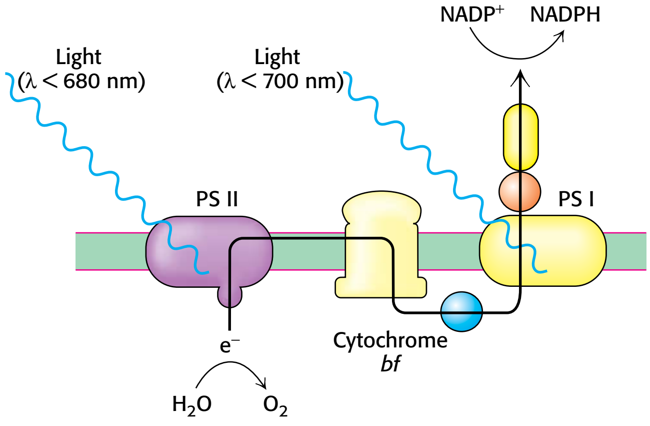


FIGURE 19.12 Two photosystems. The absorption of photons by two distinct photosystems (PS I and PS II) is required for complete electron flow from water to $NADP^+$.

Photosystem I, which responds to light with wavelengths shorter than 700 nm, uses light-derived high-energy electrons to create biosynthetic reducing power in the form of NADPH, a versatile reagent for driving biosynthetic processes. The electrons for creating one molecule of NADPH are taken from two molecules of water by photosystem II, which responds to wavelengths shorter than 680 nm. A molecule of O_2 is generated as a side product of the actions of photosystem II. The electrons travel from photosystem II to photosystem I through cytochrome *bf,* a membrane-bound complex homologous to Complex III in oxidative phosphorylation. Cytochrome *bf* generates a proton gradient across the thylakoid membrane that drives the formation of ATP. Thus, the two photosystems cooperate to produce NADPH and ATP.

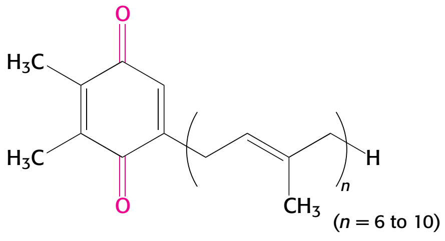
Photosystem II transfers electrons from water to plastoquinone and generates a proton gradient

Photosystem II, an enormous transmembrane assembly of more than 20 subunits, catalyzes the light-driven transfer of electrons from water to plastoquinone. This electron acceptor closely resembles ubiquinone, a component of the mitochondrial electron-transport chain. Plastoquinone cycles between an oxidized form (Q) and a reduced form (QH_2, plastoquinol). The overall reaction catalyzed by photosystem II is

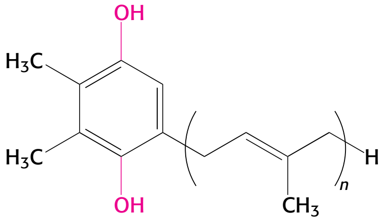
$$2\,Q + 2\,H_2O \xrightarrow{\text{Light}} O_2 + 2\,QH_2$$

The electrons in QH_2 are at a higher redox potential than those in H_2O. Recall that, in oxidative phosphorylation, electrons flow from ubiquinol to an acceptor, O_2, which is at a *lower* potential. Photosystem II drives the reaction in a thermodynamically uphill direction by using the free energy of light.

This reaction is similar to one catalyzed by the bacterial system in that a quinone is converted from its oxidized into its reduced form. Photosystem II is reasonably similar to the bacterial reaction center (Figure 19.13). The core of the photosystem is formed by D1 and D2, a pair of similar 32-kDa subunits that span the thylakoid membrane. These subunits are homologous to the L and M chains of the bacterial reaction center. Unlike the bacterial system, photosystem II contains a large number of additional subunits that bind more than 30 chlorophyll molecules altogether and increase the efficiency with which light energy is absorbed and transferred to the reaction center (Section 19.5).



Plastoquinone
(oxidized form, Q)



Plastoquinol
(reduced form, QH_2)

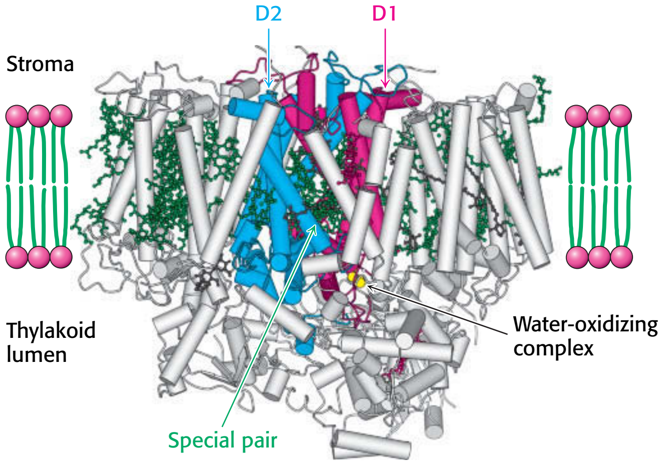


FIGURE 19.13 The structure of photosystem II. The D_1 (red) and D_2 (blue) subunits and the numerous bound chlorophyll molecules (green). *Notice* that the special pair and the water-oxidizing complex lay toward the thylakoid-lumen side of the membrane. [Drawn from 1S5L.pdb.]

The photochemistry of photosystem II begins with excitation of a special pair of chlorophyll molecules that are bound by the D1 and D2 subunits (Figure 19.14). Because the chlorophyll *a* molecules of the special pair absorb light at 680 nm, the special pair is often called *P680*. On excitation, P680 rapidly transfers an electron to a nearby pheophytin. From there, the electron is transferred first to a tightly bound plastoquinone at site Q_A and then to a mobile plastoquinone at site Q_B. This electron flow is entirely analogous to that in the bacterial system. With the arrival of a second electron and the uptake of two protons, the mobile plastoquinone is reduced to QH_2. At this point, the energy of two photons has been safely and efficiently stored in the reducing potential of QH_2.

The major difference between the bacterial system and photosystem II is the source of the electrons that are used to neutralize the positive charge formed on the special pair. *$P680^+$, a very strong oxidant, extracts electrons from water*

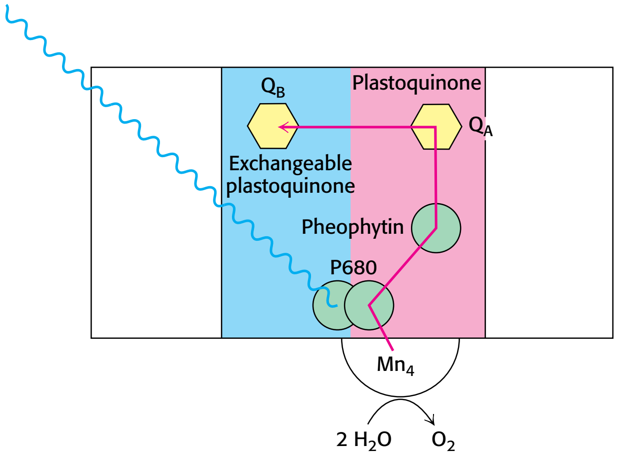


FIGURE 19.14 Electron flow through photosystem II. Light absorption induces electron transfer from P680 down an electron-transfer pathway to an exchangeable plastoquinone. The positive charge on P680 is neutralized by electron flow from water molecules bound at the manganese center.

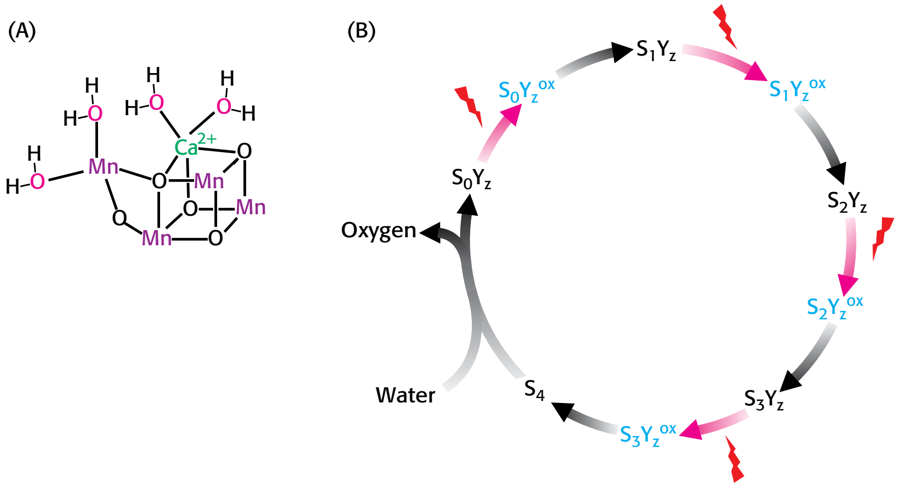
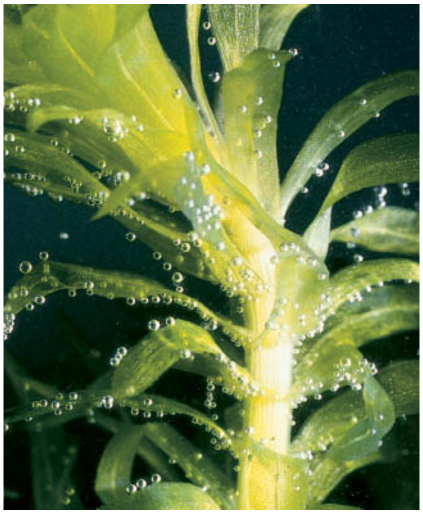


FIGURE 19.15 The core of the water-oxidizing complex. (A) The deduced core structure of the water-oxidizing complex (WOC), including four manganese ions and one calcium ion, is shown. The valence states of the individual manganese ions are not indicated because of uncertainty about the charge on the individual ions. The center is oxidized, one electron at a time, until two H_2O molecules are oxidized to form a molecule of O_2, which is then released from the complex. (B) The absorption of photons by the reaction center generates a tyrosine radical (red arrows), which then extracts electrons from the manganese ions. The structures are designated S_0 to S_4 to indicate the number of electrons that have been removed.

molecules bound at the water-oxidizing complex (WOC), also called the *manganese center.* The core of this complex includes a calcium ion, four manganese ions, and four water molecules (Figure 19.15A). Manganese was apparently evolutionarily selected for this role because of its ability to exist in multiple oxidation states and to form strong bonds with oxygen-containing species. In its reduced form, the WOC oxidizes two molecules of water to form a single molecule of oxygen. Each time the absorbance of a photon powers the removal of an electron from P680, the positively charged special pair extracts an electron from a tyrosine residue (often designated Z) of subunit D1 of the WOC, forming a tyrosine radical (Figure 19.15B). The tyrosine radical then removes an electron from a manganese ion. This process occurs four times, with the result that H_2O is oxidized to generate O_2 and H^+. Four photons must be absorbed to extract four electrons from a water molecule (Figure 19.16). The four electrons harvested from water are used to reduce two molecules of Q to QH_2. All oxygenic phototrophs, the most common type of photosynthetic organism, use the same inorganic core and protein components for capturing the energy of sunlight. A single solution to the biochemical problem of extracting electrons from water evolved billions of years ago, and has been conserved for use under a wide variety of phylogenetic and ecological circumstances.

Photosystem II spans the thylakoid membrane such that the site of quinone reduction is on the side of the stroma, whereas WOC lies in the



Evolution of oxygen is evident by the generation of bubbles in the aquatic plant *Elodea*. [Colin Milkins/Oxford Scientific Films/Photolibrary.]

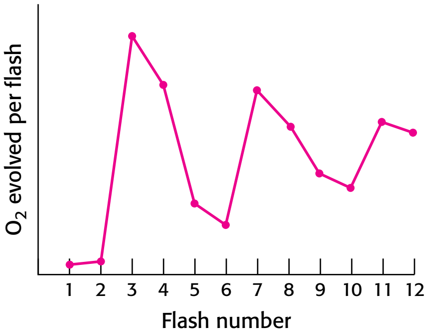


FIGURE 19.16 Four photons are required to generate one oxygen molecule. When dark-adapted chloroplasts are exposed to a brief flash of light, one electron passes through photosystem II. Monitoring the O_2 released after each flash reveals that four flashes are required to generate each O_2 molecule. The peaks in O_2 release are after the 3rd, 7th, and 11th flashes because the dark-adapted chloroplasts start in the S_1 state—that is, the one-electron reduced state.

thylakoid lumen. Thus, the two protons that are taken up with the reduction of Q to QH_2 come from the stroma, and the four protons that are liberated in the course of water oxidation are released into the lumen. This distribution of protons generates a proton gradient across the thylakoid membrane characterized by an excess of protons in the thylakoid lumen compared with the stroma (Figure 19.17).

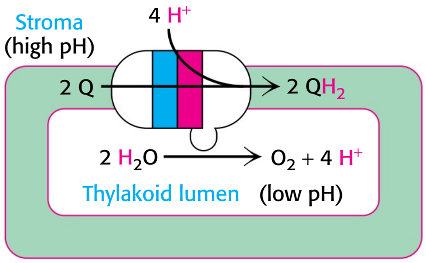


FIGURE 19.17 Proton-gradient direction. Photosystem II releases protons into the thylakoid lumen and takes them up from the stroma. The result is a pH gradient across the thylakoid membrane with an excess of protons (low pH) inside.

Cytochrome *bf* links photosystem II to photosystem I

Electrons flow from photosystem II to photosystem I through the *cytochrome* bf complex. This complex catalyzes the transfer of electrons from plastoquinol (QH_2) to plastocyanin (Pc), a small, soluble copper protein in the thylakoid lumen.

$$QH_2 + 2\,Pc(Cu^{2+}) \rightarrow Q + 2\,Pc(Cu^{+}) + 2\,H^+_{\text{thylakoid lumen}}$$

The two protons from plastoquinol are released into the thylakoid lumen. This reaction is reminiscent of that catalyzed by Complex III in oxidative phosphorylation, and most components of the cytochrome *bf* complex are homologous to those of Complex III. The cytochrome *bf* complex includes four subunits: a 23-kDa cytochrome with two *b*-type hemes, a 20-kDa Rieske-type Fe–S protein (p. 534), a 33-kDa cytochrome *f* with a *c*-type cytochrome, and a 17-kDa chain.

This complex catalyzes the reaction by proceeding through the Q cycle (Figure 18.12). In the first half of the Q cycle, plastoquinol (QH_2) is oxidized to plastoquinone (Q), one electron at a time. The electrons from plastoquinol flow through the Fe–S protein to convert oxidized plastocyanin (Pc) into its reduced form.

In the second half of the Q cycle, cytochrome *bf* reduces a molecule of plastoquinone from the Q pool to plastoquinol, taking up two protons from one side of the membrane, and then reoxidizes plastoquinol to release these protons on the other side. The enzyme is oriented so that protons are released into the thylakoid lumen and taken up from the stroma, contributing further to the proton gradient across the thylakoid membrane (Figure 19.18).

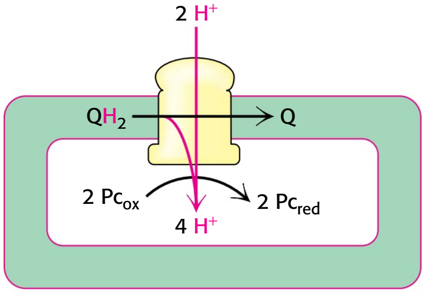


FIGURE 19.18 Cytochrome *bf* contribution to proton gradient. The cytochrome *bf* complex oxidizes QH_2 to Q through the Q cycle. Four protons are released into the thylakoid lumen in each cycle.

Photosystem I uses light energy to generate reduced ferredoxin, a powerful reductant

The final stage of the light reactions is catalyzed by photosystem I, a transmembrane complex consisting of about 15 polypeptide chains and multiple associated proteins and cofactors (Figure 19.19). The core of this system is a

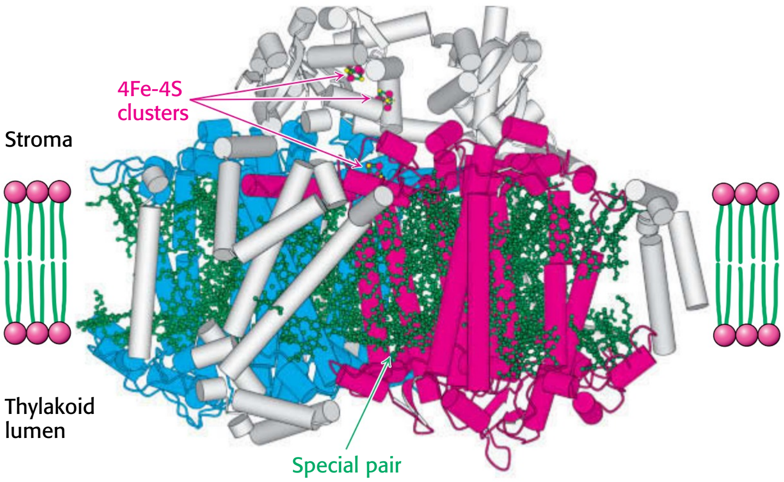


FIGURE 19.19 The structure of photosystem I. The psaA and psaB subunits are shown in red and blue, respectively. *Notice* the numerous bound chlorophyll molecules, shown in green, including the special pair, as well as the iron–sulfur clusters that facilitate electron transfer from the stroma. [Drawn from 1JBO.pdb.]

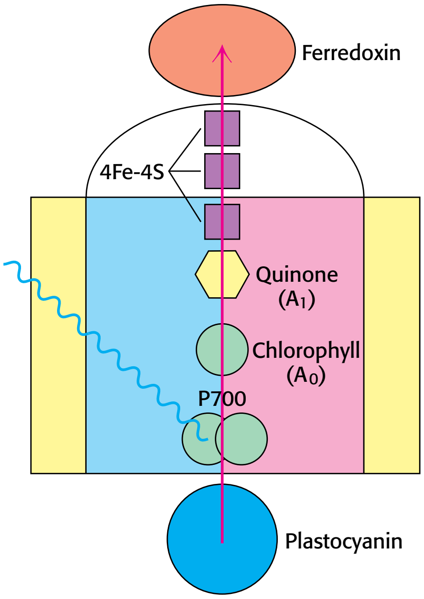


FIGURE 19.20 Electron flow through photosystem I to ferredoxin. Light absorption induces electron transfer from P700 down an electron-transfer pathway that includes a chlorophyll molecule, a quinone molecule, and three 4Fe–4S clusters to reach ferredoxin. The positive charge left on P700 is neutralized by electron transfer from reduced plastocyanin.

pair of similar subunits, psaA (83 kDa, red) and psaB (82 kDa, blue), which bind 80 chlorophyll molecules as well as other redox factors. These subunits are quite a bit larger than the core subunits of photosystem II and the bacterial reaction center. Nonetheless, they appear to be homologous; the terminal 40% of each subunit is similar to a corresponding subunit of photosystem II. A special pair of chlorophyll *a* molecules lie at the center of the structure and absorb light maximally at 700 nm. This center, called *P700*, initiates photoinduced charge separation (Figure 19.20). The electron travels from P700 down a pathway through chlorophyll at site A_0 and quinone at site A_1 to a set of 4Fe–4S clusters. The next step is the transfer of the electron to ferredoxin (Fd), a soluble protein containing a 2Fe–2S cluster coordinated to four cysteine residues (Figure 19.21). Ferredoxin transfers electrons to $NADP^+$. Meanwhile, $P700^+$ captures an electron from reduced plastocyanin provided by photosystem II to return to P700 so that P700 can be excited again. Thus, the overall reaction catalyzed by photosystem I is a simple one-electron oxidation–reduction reaction.

$$\mathrm{Pc(Cu^+) + Fd_{ox}} \xrightarrow{\text{Light}} \mathrm{Pc(Cu^{2+}) + Fd_{red}}$$

Given that the reduction potentials for plastocyanin and ferredoxin are +0.37 V and −0.45 V, respectively, the standard free energy for this reaction is +79.1 kJ mol^{-1} (+18.9 kcal mol^{-1}). This uphill reaction is driven by the absorption of a 700-nm photon, which has an energy of 171 kJ mol^{-1} (40.9 kcal mol^{-1}).

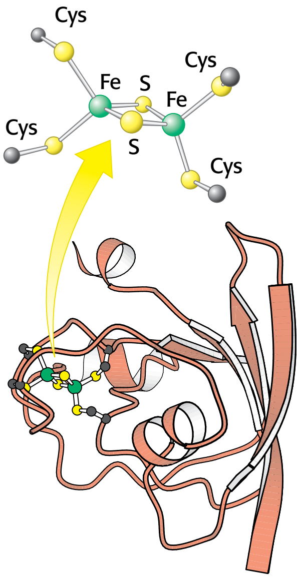


FIGURE 19.21 Structure of ferredoxin. In plants, ferredoxin contains a 2Fe-2S cluster. This protein accepts electrons from photosystem I and carries them to ferredoxin–$NADP^+$ reductase. [Drawn from 1FXA.pdb.]

Ferredoxin–$NADP^+$ reductase converts $NADP^+$ into NADPH

The reduced ferredoxin generated by photosystem I is a strong reductant that is used as an electron source for a variety of reactions, most notably the fixation of N_2 into NH_3 (Section 24.1). However, ferredoxin is not useful for driving many reactions, in part because ferredoxin carries only one available electron. In contrast, NADPH, a two-electron reductant, is a widely used electron donor in biosynthetic processes, including the reactions of the Calvin cycle (Chapter 20). How is reduced ferredoxin used to drive the reduction of $NADP^+$ to NADPH? This reaction is catalyzed by *ferredoxin–$NADP^+$ reductase,* a flavoprotein with an FAD prosthetic

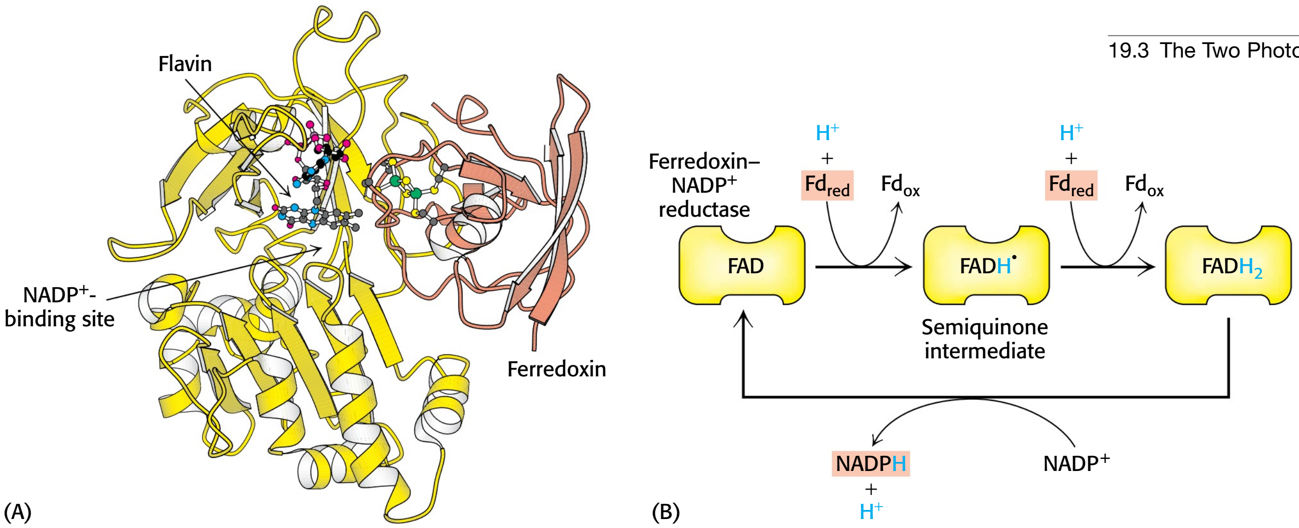

Flavin
NADP+-binding site
Ferredoxin
(A)
Ferredoxin–NADP+ reductase
H+
+
Fdred
Fdox
FAD
FADH•
Semiquinone intermediate
FADH2
NADPH
+
H+
NADP+
(B)

FIGURE 19.22 Structure and function of ferredoxin–NADP$^+$ reductase. (A) Structure of ferredoxin–NADP$^+$ reductase. This enzyme accepts electrons, one at a time, from ferredoxin (shown in orange). (B) Ferredoxin–NADP$^+$ reductase first accepts two electrons, generated by photosystem I, and two protons from the lumen to form two molecules of reduced ferredoxin (Fd). Reduced ferredoxin is then used to form $FADH_2$, which then transfers two electrons and a proton to NADP$^+$ to form NADPH in the stroma. [Drawn from 1EWY.pdb.]

group (Figure 19.22A). The bound FAD moiety accepts two electrons and two protons from two molecules of reduced ferredoxin to form $FADH_2$ (Figure 19.22B). The enzyme then transfers a hydride ion (H^-) to NADP$^+$ to form NADPH. This reaction takes place on the stromal side of the membrane. Hence, the uptake of a proton in the reduction of NADP$^+$ further contributes to the generation of the proton gradient across the thylakoid membrane.

The cooperation between photosystem I and photosystem II creates a flow of electrons from H_2O to NADP$^+$. The pathway of electron flow is called the *Z scheme of photosynthesis* because the redox diagram from P680 to P700* looks like the letter Z (Figure 19.23).

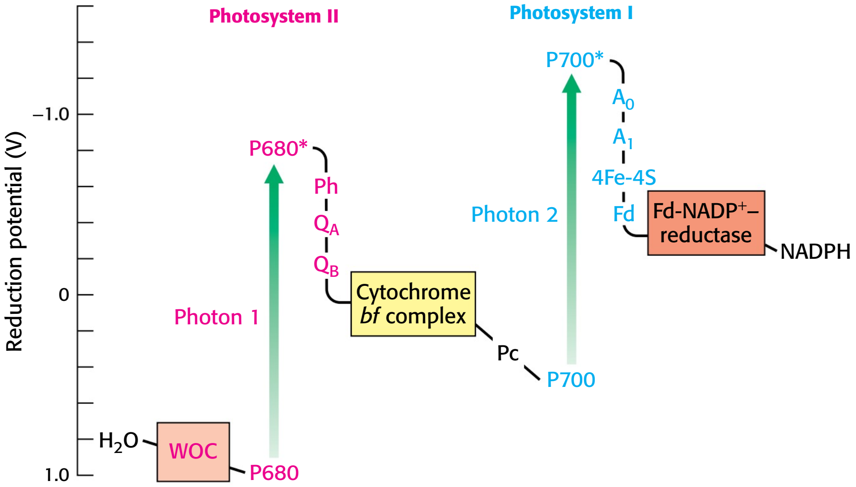
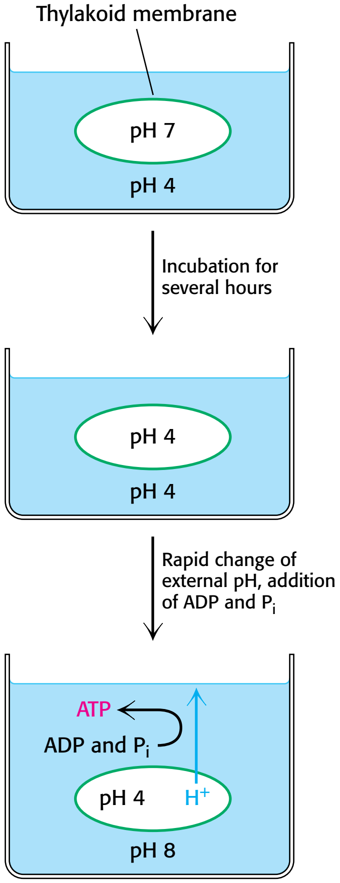

Reduction potential (V)
−1.0
0
1.0
Photosystem II
Photosystem I
P680*
Ph
QA
QB
Cytochrome bf complex
Pc
P700
Photon 1
H2O
WOC
P680
P700*
A0
A1
4Fe-4S
Fd
Fd-NADP+–reductase
NADPH
Photon 2

FIGURE 19.23 Pathway of electron flow from H_2O to NADP$^+$ in photosynthesis. This endergonic reaction is made possible by the absorption of light by photosystem II (P680) and photosystem I (P700). Abbreviations: Ph, pheophytin; Q_A and Q_B, plastoquinone-binding proteins; Pc, plastocyanin; A_0 and A_1, acceptors of electrons from P700*; Fd, ferredoxin; WOC, water-oxidizing complex.



Thylakoid membrane
pH 7
pH 4
Incubation for several hours
pH 4
pH 4
Rapid change of external pH, addition of ADP and P_i
ATP
ADP and P_i
pH 4
H^+
pH 8

FIGURE 19.24 Jagendorf's demonstration. Chloroplasts synthesize ATP after the imposition of a pH gradient.

19.4 A Proton Gradient across the Thylakoid Membrane Drives ATP Synthesis

In 1966, André Jagendorf showed that chloroplasts synthesize ATP in the dark when an artificial pH gradient is imposed across the thylakoid membrane. To create this transient pH gradient, he soaked chloroplasts in a pH 4 buffer for several hours and then rapidly mixed them with a pH 8 buffer containing ADP and P_i. The pH of the stroma suddenly increased to 8, whereas the pH of the thylakoid space remained at 4. *A burst of ATP synthesis then accompanied the disappearance of the pH gradient across the thylakoid membrane* (Figure 19.24). This incisive experiment was one of the first to unequivocally support the hypothesis put forth by Peter Mitchell that ATP synthesis is driven by proton-motive force (Section 18.4).

The principles of ATP synthesis in chloroplasts are nearly identical with those in mitochondria. *ATP formation is driven by a proton-motive force in both photophosphorylation and oxidative phosphorylation.* We have seen how light induces electron transfer through photosystems II and I and the cytochrome *bf* complex. At various stages in this process, protons are released into the thylakoid lumen or taken up from the stroma, generating a proton gradient. The gradient is maintained because the thylakoid membrane is essentially impermeable to protons. *The thylakoid space becomes markedly acidic, with the pH approaching 4. The light-induced transmembrane proton gradient is about 3.5 pH units.* As discussed in Section 18.4, energy inherent in the proton gradient, called the *proton-motive force* (Δp), is described as the sum of two components: a charge gradient and a chemical gradient. In chloroplasts, nearly all of Δp arises from the pH gradient, whereas in mitochondria, the contribution from the membrane potential is larger. The reason for this difference is that the thylakoid membrane is quite permeable to Cl^- and Mg^{2+}. The light-induced transfer of H^+ into the thylakoid space is accompanied by the transfer of either Cl^- in the same direction or Mg^{2+} (1 Mg^{2+} per 2 H^+) in the opposite direction. Consequently, electrical neutrality is maintained and no membrane potential is generated. The influx of Mg^{2+} into the stroma plays a role in the regulation of the Calvin Cycle (Section 20.2). A pH gradient of 3.5 units across the thylakoid membrane corresponds to a proton-motive force of 0.20 V or a ΔG of -20.0 kJ mol^{-1} (-4.8 kcal mol^{-1}).

The ATP synthase of chloroplasts closely resembles those of mitochondria and prokaryotes

The proton-motive force generated by the light reactions is converted into ATP by the *ATP synthase* of chloroplasts, also called the *CF_1–CF_0 complex* (*C* stands for chloroplast; *F* for factor). CF_1–CF_0 ATP synthase closely resembles the F_1–F_0 complex of mitochondria (Section 18.4). CF_0 conducts protons across the thylakoid membrane, whereas CF_1 catalyzes the formation of ATP from ADP and P_i.

CF_0 is embedded in the thylakoid membrane. It consists of four different polypeptide chains known as I (17 kDa), II (16.5 kDa), III (8 kDa), and IV (27 kDa) having an estimated stoichiometry of 1 : 2 : 10–14 : 1. Subunits I and II have sequence similarity to subunit **b** of the mitochondrial F_0 subunit, III corresponds to subunit **c** of the mitochondrial complex, and subunit IV is similar in sequence to subunit **a.** CF_1, the site of ATP synthesis, has a subunit composition $\alpha_3\beta_3\gamma\delta\varepsilon$. The β subunits contain the catalytic sites, similarly to the F_1 subunit of mitochondrial ATP synthase. Remarkably, the β subunits of ATP synthase in corn chloroplasts are more

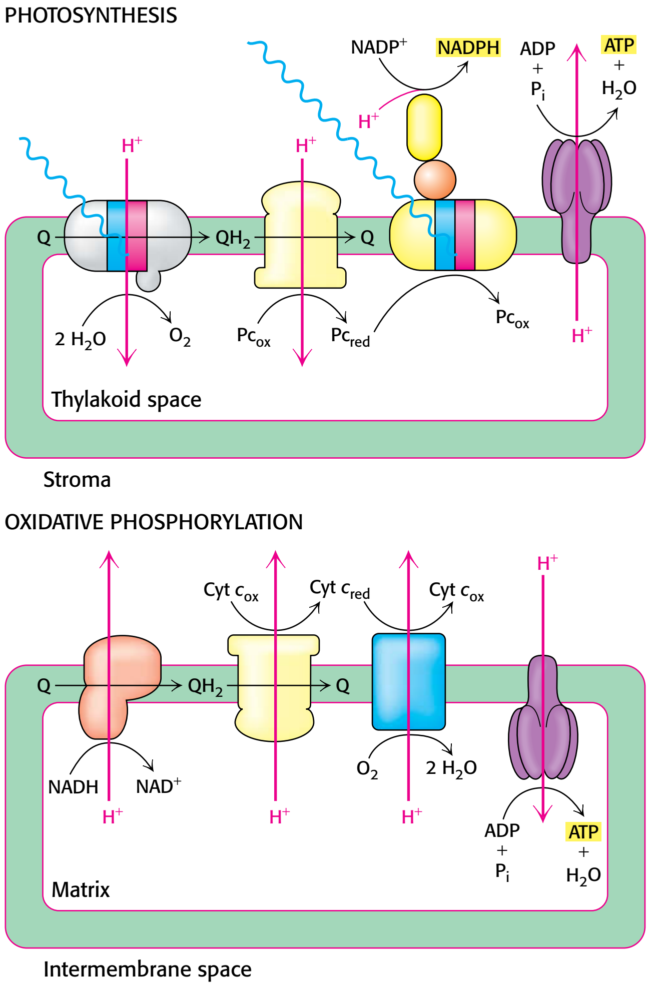

PHOTOSYNTHESIS
NADP$^+$
NADPH
ADP + P_i
ATP + H_2O
H^+
H^+
H^+
Q
QH_2
Q
2 H_2O
O_2
Pc_{ox}
Pc_{red}
Pc_{ox}
H^+
Thylakoid space
Stroma
OXIDATIVE PHOSPHORYLATION
H^+
Cyt c_{ox}
Cyt c_{red}
Cyt c_{ox}
Q
QH_2
Q
O_2
2 H_2O
NADH
NAD$^+$
H^+
H^+
H^+
ADP + P_i
ATP + H_2O
Matrix
Intermembrane space

FIGURE 19.25 Comparison of photosynthesis and oxidative phosphorylation. The light-induced electron transfer in photosynthesis drives protons into the thylakoid lumen. The excess protons flow out of the lumen through ATP synthase to generate ATP in the stroma. In oxidative phosphorylation, electron flow down the electron-transport chain pumps protons out of the mitochondrial matrix. Excess protons from the intermembrane space flow into the matrix through ATP synthase to generate ATP in the matrix.

than 60% identical in amino acid sequence with those of human ATP synthase, despite the passage of approximately 1 billion years since the separation of the plant and animal kingdoms.

Note that the membrane orientation of CF_1–CF_0 is reversed compared with that of the mitochondrial ATP synthase (Figure 19.25). However, the functional orientation of the two synthases is identical: protons flow from the lumen through the enzyme to the stroma or matrix where ATP is synthesized. Because CF_1 is on the stromal surface of the thylakoid membrane, the newly synthesized ATP is released directly into the stromal space. Likewise, NADPH formed by photosystem I is released into the stromal space. Thus, *ATP and NADPH, the products of the light reactions of photosynthesis, are appropriately positioned for the subsequent dark reactions, in which CO_2 is converted into carbohydrate.*

The activity of chloroplast ATP synthase is regulated

The activity of the ATP synthase is sensitive to the redox conditions in the chloroplast. For maximal activity, a specific disulfide bond in the γ subunit must be reduced to two cysteines. The reductant is reduced thioredoxin, which is formed from ferredoxin generated in photosystem I, by ferredoxin–thioredoxin reductase, an iron–sulfur containing enzyme.

$$2\ \text{Reduced ferredoxin} + \text{thioredoxin disulfide} \xrightleftharpoons{\text{Ferredoxin-thioredoxin reductase}} 2\ \text{oxidized ferredoxin} + \text{reduced thioredoxin} + 2\text{H}^+$$

Conformational changes in the ε subunit also contribute to synthase regulation. The ε subunit appears to exist in two conformations. One conformation inhibits ATP hydrolysis by the synthase, while the other, which is generated by an increase in the proton-motive force, allows ATP synthesis and facilitates the reduction of the disulfide bond in the γ subunit. Thus, synthase activity is maximal when biosynthetic reducing power and a proton gradient are available. We will see in Chapter 20 that redox regulation is also important in photosynthetic carbon metabolism.

Cyclic electron flow through photosystem I leads to the production of ATP instead of NADPH

On occasion, when the ratio of NADPH to $NADP^+$ is very high, as might be the case if there was another source of electrons to form NADPH (Section 20.3), $NADP^+$ may be unavailable to accept electrons from reduced ferredoxin. In these circumstances, specific large protein complexes allow cyclic electron flow that powers ATP synthesis. Electrons arising from P700, the reaction center of photosystem I, generate reduced ferredoxin. The electron in reduced ferredoxin is transferred to the cytochrome *bf* complex rather than to $NADP^+$. This electron then flows back through the cytochrome *bf* complex to reduce plastocyanin, which can then be reoxidized by $P700^+$ to complete a cycle. The net outcome of this cyclic flow of electrons is the pumping of protons by the cytochrome *bf* complex. The resulting proton gradient then drives the synthesis of ATP. In this process, called *cyclic photophosphorylation, ATP is generated without the concomitant formation of NADPH* (Figure 19.26). Photosystem II does not participate in cyclic photophosphorylation, and so O_2 is not formed from H_2O.

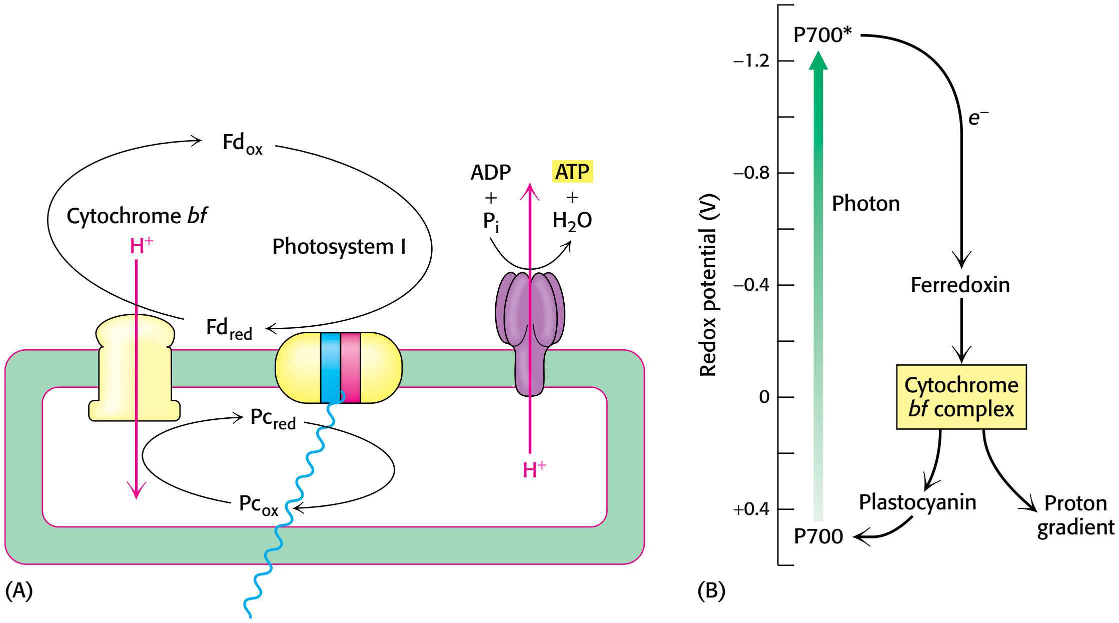


FIGURE 19.26 Cyclic photophosphorylation. (A) In this pathway, electrons from reduced ferredoxin are transferred to cytochrome *bf* rather than to ferredoxin–$NADP^+$ reductase. The flow of electrons through cytochrome *bf* pumps protons into the thylakoid lumen. These protons flow through ATP synthase to generate ATP. Neither NADPH nor O_2 is generated by this pathway. (B) A scheme showing the energetic basis for cyclic photophosphorylation. Abbreviations: Fd, ferredoxin; Pc, plastocyanin.

The absorption of eight photons yields one O_2, two NADPH, and three ATP molecules

We can now estimate the overall stoichiometry for the light reactions. The absorption of four photons by photosystem II generates one molecule of O_2 and releases four protons into the thylakoid lumen. The two molecules of plastoquinol are oxidized by the Q cycle of the cytochrome *bf* complex to release eight protons into the lumen. Finally, the electrons from four molecules of reduced plastocyanin are driven to ferredoxin by the absorption of four additional photons. The four molecules of reduced ferredoxin generate two molecules of NADPH. Thus, the overall reaction is

$$2\,H_2O + 2\,NADP^+ + 10\,H^+_{stroma} \rightarrow O_2 + 2\,NADPH + 12\,H^+_{lumen}$$

The 12 protons released in the lumen can then flow through ATP synthase. Let us assume that there are 12 subunit III components in CF_0. We expect that 12 protons must pass through CF_0 to complete one full rotation of CF_1. A single rotation generates three molecules of ATP. Given the ratio of 3 ATP for 12 protons, the overall reaction is

$$\begin{array}{r}
2\,H_2O + 2\,NADP^+ + 10\,H^+_{stroma} \longrightarrow O_2 + 2\,NADPH + 12\,H^+_{lumen} \\
\underline{3\,ADP^{3-} + 3\,P_i^{2-} + 3\,H^+ + 12\,H^+_{lumen} \longrightarrow 3\,ATP^{4-} + 3\,H_2O + 12\,H^+_{stroma}} \\
2\,NADP^+ + 3\,ADP^{3-} + 3\,P_i^{2-} + H^+ \longrightarrow O_2 + 2\,NADPH + 3\,ATP^{4-} + H_2O
\end{array}$$

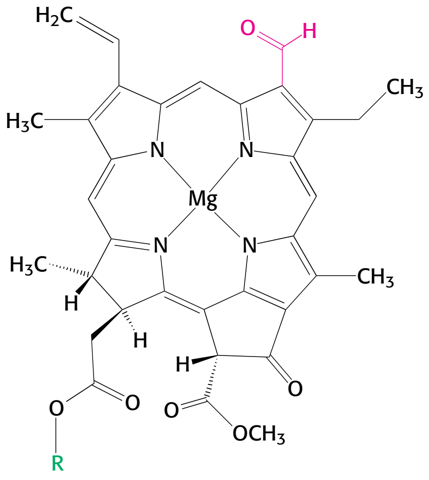
Thus, eight photons are required to yield three molecules of ATP (2.7 photons/ATP).

Cyclic photophosphorylation is a somewhat more productive way to synthesize ATP than noncyclic photophosphorylation (the Z scheme). The absorption of four photons by photosystem I leads to the release of eight protons into the lumen by the cytochrome *bf* system. These protons flow through ATP synthase to yield two molecules of ATP. Thus, each two absorbed photons yield one molecule of ATP. No NADPH is produced.

19.5 Accessory Pigments Funnel Energy into Reaction Centers

A light-harvesting system that relied only on the chlorophyll *a* molecules of the special pair would be rather inefficient for two reasons. First, chlorophyll *a* molecules absorb light only at specific wavelengths (Figure 19.6). A large gap is present in the middle of the visible region between approximately 450 and 650 nm. This gap falls right at the peak of the solar spectrum, and so failure to collect this light would constitute a considerable lost opportunity. Second, even on a cloudless day, many photons that can be absorbed by chlorophyll *a* pass through the chloroplast without being absorbed, because the density of chlorophyll *a* molecules in a reaction center is not very great. Accessory pigments, both additional chlorophylls and other classes of molecules, are closely associated with reaction centers. *These pigments absorb light and funnel the energy to the reaction center for conversion into chemical forms*. Accessory pigments prevent the reaction center from sitting idle.

Chlorophyll b and *carotenoids* are important accessory pigments that funnel energy to the reaction center. Chlorophyll *b* differs from chlorophyll *a* in having a formyl group in place of a methyl group. This small difference shifts its two major absorption peaks toward the center of the visible region.



Chlorophyll *b*

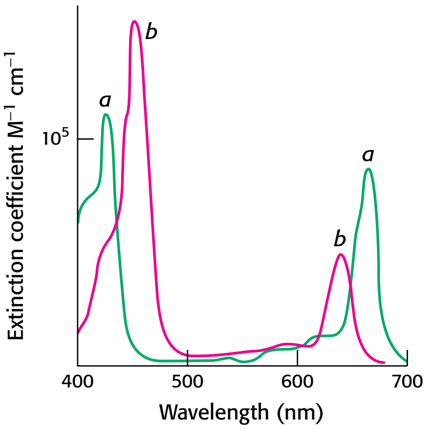
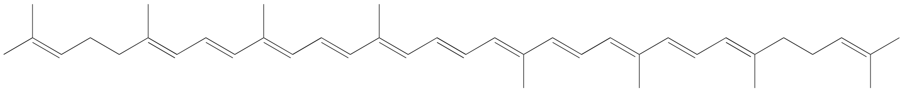


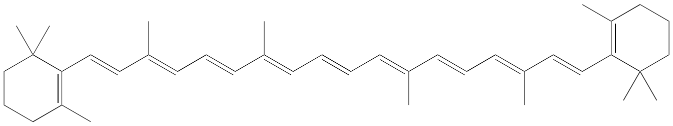
FIGURE 19.27 Absorption spectra of chlorophylls *a* and *b*.

In particular, chlorophyll *b* efficiently absorbs light with wavelengths between 450 and 500 nm (Figure 19.27).

Carotenoids are extended polyenes that absorb light between 400 and 500 nm. The carotenoids are responsible for most of the yellow and red colors of fruits and flowers, and they provide the brilliance of fall, when the chlorophyll molecules degrade, revealing the carotenoids. Tomatoes are especially rich in lycopene, which accounts for their red color, whereas carrots and pumpkins have abundant β-carotene, accounting for their orange color.



Lycopene



β-Carotene

In addition to their role in transferring energy to reaction centers, the carotenoids serve a safeguarding function. Carotenoids suppress damaging photochemical reactions, particularly those including oxygen, that can be induced by bright sunlight. This protection may be especially important in the fall, when the primary pigment chlorophyll is being degraded and thus not able to absorb light energy. Plants lacking carotenoids are quickly killed on exposure to light and oxygen.

Resonance energy transfer allows energy to move from the site of initial absorbance to the reaction center

How is energy funneled from accessory pigments to a reaction center? The absorption of a photon does not always lead to electron excitation and transfer. More commonly, excitation energy is transferred from one molecule to a nearby molecule through electromagnetic interactions through space (Figure 19.28). The rate of this process, called *resonance energy transfer,* depends strongly on the distance between the energy-donor and the energy-acceptor molecules; an increase in the distance between the donor and the acceptor by a factor of 2 typically results in a decrease in the energy-transfer rate by a factor of $2^6 = 64$. For reasons of conservation of energy, energy

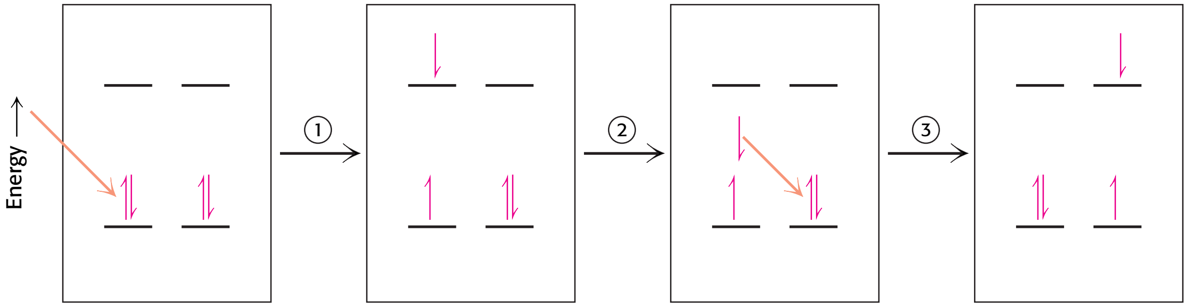


FIGURE 19.28 Resonance energy transfer. (1) An electron can accept energy from electromagnetic radiation of appropriate wavelength and jump to a higher energy state. (2) When the excited electron falls back to its lower energy state, the absorbed energy is released. (3) The released energy can be absorbed by an electron in a nearby molecule, and this electron jumps to a high energy state.

transfer must be from a donor in the excited state to an acceptor of equal or lower energy. *The excited state of the special pair of chlorophyll molecules is lower in energy than that of single chlorophyll molecules, allowing reaction centers to trap the energy transferred from other molecules.*

What is the structural relationship between the accessory pigment and the reaction center? The accessory pigments are arranged in numerous *light-harvesting complexes* that completely surround the reaction center (Figure 19.29). The 26-kDa subunit of light-harvesting complex II (LHC-II) is the most abundant membrane protein in chloroplasts. This subunit binds seven chlorophyll *a* molecules, six chlorophyll *b* molecules, and two carotenoid molecules. Similar light-harvesting assemblies exist in photosynthetic bacteria (Figure 19.30).

The components of photosynthesis are highly organized

The intricacy of photosynthesis, seen already in the elaborate interplay of complex components, extends even to the placement of the components in the thylakoid membranes. *Thylakoid membranes of most plants are differentiated into stacked* (appressed) *and unstacked* (nonappressed) *regions* (Figures 19.2 and 19.3). Stacking increases the amount of thylakoid membrane in a given chloroplast volume. Both regions surround a common internal thylakoid space, but only unstacked regions make direct contact with the chloroplast stroma. Stacked and unstacked regions differ in the nature of their photosynthetic assemblies (Figure 19.31). Photosystem I and ATP synthase are located almost exclusively in unstacked regions, whereas photosystem II is present mostly in stacked regions. The cytochrome *bf* complex is found in both regions. Plastoquinone and plastocyanin are the mobile carriers of electrons between assemblies located in different regions of the thylakoid membrane. A common internal thylakoid space enables protons liberated by photosystem II in stacked membranes to be utilized by ATP synthase molecules that are located far away in unstacked membranes.

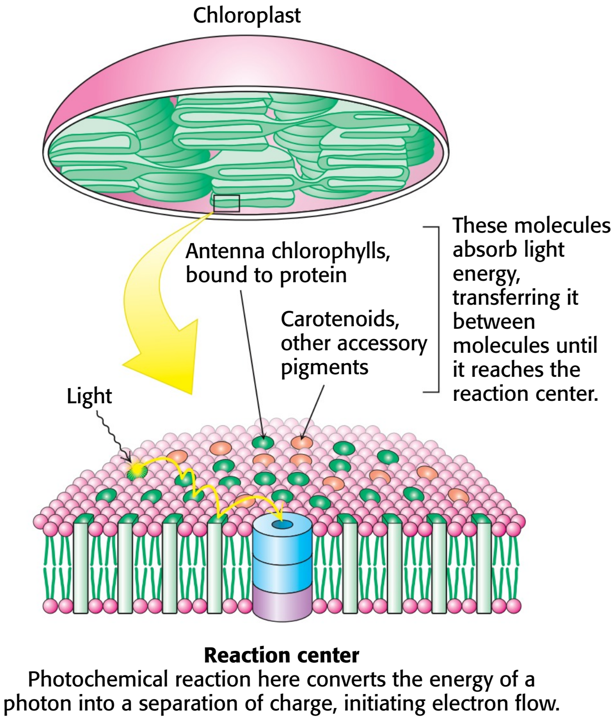

Chloroplast
Antenna chlorophylls, bound to protein
Carotenoids, other accessory pigments
These molecules absorb light energy, transferring it between molecules until it reaches the reaction center.
Light
Reaction center
Photochemical reaction here converts the energy of a photon into a separation of charge, initiating electron flow.

FIGURE 19.29 Energy transfer from accessory pigments to reaction centers. Light energy absorbed by accessory chlorophyll molecules or other pigments can be transferred to reaction centers by resonance energy transfer (yellow arrows), where it drives photoinduced charge separation. [Source: D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 5th ed. (W. H. Freeman and Company, 2008), Fig. 19.52.]

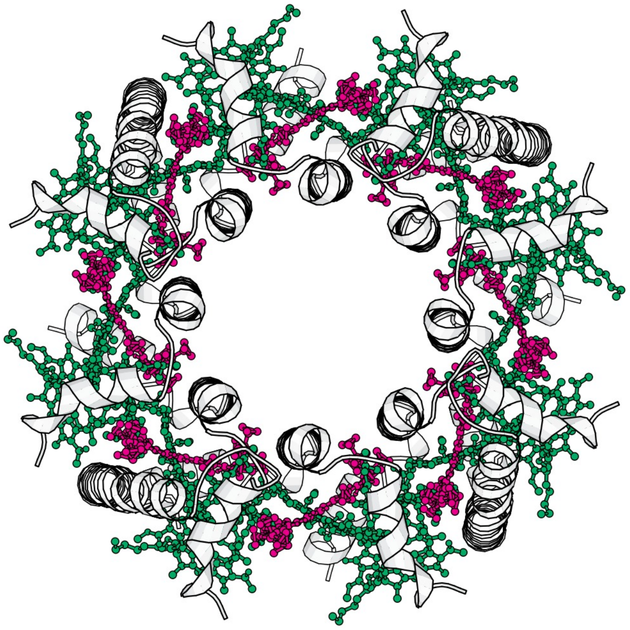


FIGURE 19.30 Structure of a bacterial light-harvesting complex. Eight polypeptides, each of which binds three chlorophyll molecules (green) and a carotenoid molecule (red), surround a central cavity that contains the reaction center (not shown). *Notice* the high concentration of accessory pigments that surround the reaction center. [Drawn from 1LGH.pdb.]

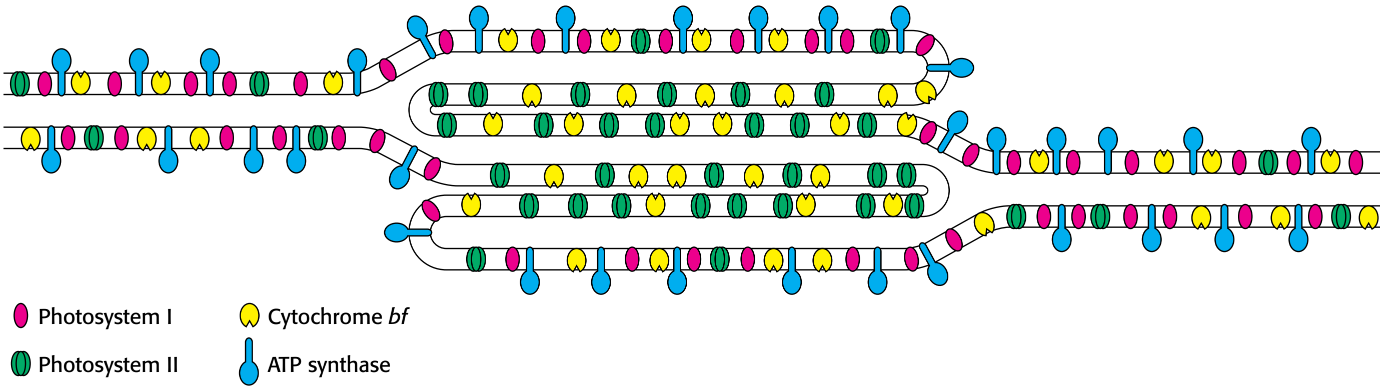
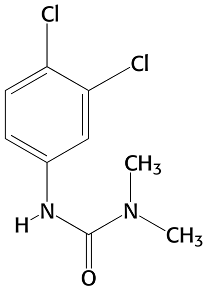
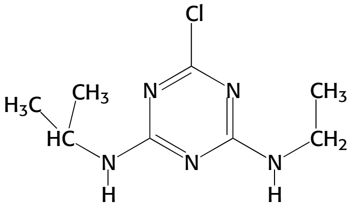


FIGURE 19.31 Location of photosynthesis components. Photosynthetic assemblies are differentially distributed in the stacked (appressed) and unstacked (nonappressed) regions of thylakoid membranes. [Information from Dr. Jan M. Anderson and Dr. Bertil Andersson.]


Diuron


Atrazine

What is the functional significance of this lateral differentiation of the thylakoid membrane system? The positioning of photosystem I in the unstacked membranes also gives it direct access to the stroma for the reduction of $NADP^+$. ATP synthase is also located in the unstacked region to provide space for its large CF_1 globule and to give access to ADP. In contrast, the tight quarters of the appressed region pose no problem for photosystem II, which interacts with a small polar electron donor (H_2O) and a highly lipid soluble electron carrier (plastoquinone).

Many herbicides inhibit the light reactions of photosynthesis

Many commercial herbicides kill weeds by interfering with the action of photosystem II or photosystem I. Inhibitors of photosystem II block electron flow, whereas inhibitors of photosystem I divert electrons from the terminal part of this photosystem. Photosystem II inhibitors include urea derivatives such as *diuron* and triazine derivatives such as *atrazine*. These chemicals bind to the Q_B site of the D1 subunit of photosystem II and block the formation of plastoquinol (QH_2).

Paraquat (1,1′-dimethyl-4-4′-bipyridinium) is an inhibitor of photosystem I. Paraquat, a dication, can accept electrons from photosystem I to become a radical. This radical reacts with O_2 to produce reactive oxygen species such as superoxide (O_2^-) and hydroxyl radical (OH•). Such reactive oxygen species react with many biomolecules, including double bonds in membrane lipids, damaging the membrane.

19.6 The Ability to Convert Light into Chemical Energy Is Ancient

The ability to convert light energy into chemical energy is a tremendous evolutionary advantage. Geological evidence suggests that oxygenic photosynthesis became important approximately 2 billion years ago. Anoxygenic photosynthetic systems arose much earlier in the 3.5-billion-year history of life on Earth (Table 19.1). The photosynthetic system of the nonsulfur purple bacterium *Rhodopseudomonas viridis* has most features common to oxygenic photosynthetic systems and clearly predates them. Green sulfur bacteria such as *Chlorobium thiosulfatophilum* carry out a reaction that also seems to have appeared before oxygenic photosynthesis and is even more similar to oxygenic photosynthesis than the photosystem of

TABLE 19.1 Major groups of photosynthetic prokaryotes

Bacteria	Photosynthetic electron donor	O_2 use
Green sulfur	H_2, H_2S, S	Anoxygenic
Green nonsulfur	Variety of amino acids and organic acids	Anoxygenic
Purple sulfur	H_2, H_2S, S	Anoxygenic
Purple nonsulfur	Usually organic molecules	Anoxygenic
Cyanobacteria	H_2O	Oxygenic

R. viridis. Reduced sulfur species such as H_2S are electron donors in the overall photosynthetic reaction

$$CO_2 + 2\,H_2S \xrightarrow{\text{Light}} (CH_2O) + 2\,S + H_2O$$

Nonetheless, photosynthesis did not evolve immediately at the origin of life. No photosynthetic organisms have been discovered in the domain of Archaea, implying that photosynthesis evolved in the domain of Bacteria after Archaea and Bacteria diverged from a common ancestor. All domains of life do have electron-transport chains in common, however. As we have seen, components such as the ubiquinone–cytochrome *c* oxidoreductase and cytochrome *bf* family are present in both respiratory and photosynthetic electron-transport chains. These components were the foundations on which light-energy-capturing systems evolved.

Artificial photosynthetic systems may provide clean, renewable energy

As we have learned, photosynthetic organisms use sunlight to oxidize H_2O, producing O_2 and protons used to power ATP synthesis and generate NADPH. Research is currently underway to try to mimic this process in order to provide clean energy. Photovoltaic cells can use light energy to oxidize water, producing O_2 as well as H_2. Hydrogen gas is a fuel that, upon reaction with oxygen, generates energy and only water as a waste product. Major difficulties in creating efficient photovoltaic cells are that the materials required are not durable and often not readily available. Recent work suggests that semiconductors composed of organic–inorganic material show great promise. One such material, perovskite ($CH_3NH_3PbI_3$), is remarkably efficient at capturing sunlight. It is humbling to realize that, although photosynthetic organisms have been oxidizing water for billions of years, researchers today struggle to replicate the process.

SUMMARY

19.1 Photosynthesis Takes Place in Chloroplasts

The proteins that participate in the light reactions of photosynthesis are located in the thylakoid membranes of chloroplasts. The light reactions result in (1) the creation of reducing power for the production of NADPH, (2) the generation of a transmembrane proton gradient for the formation of ATP, and (3) the production of O_2.

19.2 Light Absorption by Chlorophyll Induces Electron Transfer

Chlorophyll molecules absorb light quite efficiently because they are polyenes. An electron excited to a high-energy state by the absorption of a photon can move to nearby electron acceptors. In photosynthesis, an excited electron leaves a pair of associated chlorophyll molecules known as the special pair. The functional core of photosynthesis, a reaction center from a photosynthetic bacterium, has been studied in great detail.

In this system, the electron moves from the special pair (containing bacteriochlorophyll) to a bacteriopheophytin (a bacteriochlorophyll lacking the central magnesium ion) to quinones. The reduction of quinones leads to the generation of a proton gradient, which drives ATP synthesis in a manner analogous to that of oxidative phosphorylation.

19.3 Two Photosystems Generate a Proton Gradient and NADPH in Oxygenic Photosynthesis

Photosynthesis in green plants is mediated by two linked photosystems. In photosystem II, the excitation of a special pair of chlorophyll molecules called P680 leads to electron transfer to plastoquinone in a manner analogous to that of the bacterial reaction center. The electrons are replenished by the extraction of electrons from a water molecule at the water-oxidizing complex, which contains four manganese ions. One molecule of O_2 is generated at this center for each four electrons transferred. The plastoquinol produced at photosystem II is reoxidized by the cytochrome *bf* complex, which transfers the electrons to plastocyanin, a soluble copper protein. From plastocyanin, the electrons enter photosystem I. In photosystem I, the excitation of special pair P700 releases electrons that flow to ferredoxin, a powerful reductant. Ferredoxin–$NADP^+$ reductase, a flavoprotein, then catalyzes the formation of NADPH. A proton gradient is generated as electrons pass through photosystem II, through the cytochrome *bf* complex, and through ferredoxin–$NADP^+$ reductase.

19.4 A Proton Gradient across the Thylakoid Membrane Drives ATP Synthesis

The proton gradient across the thylakoid membrane creates a proton-motive force, used by ATP synthase to form ATP. The ATP synthase of chloroplasts (also called the CF_0–CF_1 complex) closely resembles the ATP-synthesizing assemblies of bacteria and mitochondria. If the NADPH:$NADP^+$ ratio is high, electrons transferred to ferredoxin by photosystem I can reenter the cytochrome *bf* complex. This process, called cyclic photophosphorylation, leads to the generation of a proton gradient by the cytochrome *bf* complex without the formation of NADPH or O_2.

19.5 Accessory Pigments Funnel Energy into Reaction Centers

Light-harvesting complexes that surround the reaction centers contain additional molecules of chlorophyll *a*, as well as carotenoids and chlorophyll *b* molecules, which absorb light in the center of the visible spectrum. These accessory pigments increase the efficiency of light capture by absorbing light and transferring the energy to reaction centers through resonance energy transfer.

19.6 The Ability to Convert Light into Chemical Energy Is Ancient

The photosystems have structural features in common that suggest a common evolutionary origin. Similarities in organization and molecular structure to those of oxidative phosphorylation suggest that the photosynthetic apparatus evolved from an early energy-transduction system.

KEY TERMS

light reactions (p. 566)
chloroplast (p. 567)
stroma (p. 567)
thylakoid (p. 567)
granum (p. 567)
chlorophyll *a* (p. 568)
photoinduced charge separation (p. 569)
reaction center (p. 569)
special pair (p. 570)
P960 (p. 571)
photosystem I (PS I) (p. 572)
photosystem II (PS II) (p. 572)
P680 (p. 573)
water-oxidizing complex (WOC) (manganese center) (p. 574)
cytochrome *bf* (p. 575)
P700 (p. 576)
Z scheme of photosynthesis (p. 577)
proton-motive force (p. 578)
ATP synthase (CF_1–CF_0 complex) (p. 578)
cyclic photophosphorylation (p. 580)
carotenoid (p. 581)
light-harvesting complex (p. 583)

PROBLEMS

1. *A crucial prereq*. Human beings do not produce energy by photosynthesis, yet this process is critical to our survival. Explain.

2. *The accounting.* What is the overall reaction for the light reactions of photosynthesis?

3. *Like a fife and drum.* Match each term with its description.

(a) Light reactions _____
(b) Chloroplasts _____
(c) Reaction center _____
(d) Chlorophyll _____
(e) Light-harvesting complex _____
(f) Photosystem I _____
(g) Photosystem II _____
(h) Cytochrome *bf* complex _____
(i) Water-oxidizing complex _____
(j) ATP synthase _____

1. Uses resonance energy transfer to reach the reaction center
2. Powers the formation of NADPH
3. Pumps protons
4. Site of photoinduced charge separation
5. Cellular location of photosynthesis
6. CF_1–CF_0 complex
7. Generates ATP, NADPH, and O_2
8. Site of oxygen generation
9. Transfers electrons from H_2O to P680
10. Primary photosynthetic pigment

4. *A single wavelength*. Photosynthesis can be measured by measuring the rate of oxygen production. When plants are exposed to light of wavelength 680 nm, more oxygen is evolved than if the plants are exposed to light of 700 nm. Explain.

5. *Combining wavelengths*. If plants described in problem 4 are illuminated by a combination of light of 680 nm and 700 nm, the oxygen production exceeds that of either wavelength alone. Explain.

6. *Complementary powers*. Photosystem I produces a powerful reductant, whereas photosystem II produces a powerful oxidant. Identify the reductant and oxidant and describe their roles.

7. *If a little is good*. What is the advantage of having an extensive set of thylakoid membranes in the chloroplasts?

8. *Cooperation*. Explain how light-harvesting complexes enhance the efficiency of photosynthesis.

9. *One thing leads to another*. What is the ultimate electron acceptor in photosynthesis? The ultimate electron donor? What powers the electron flow between the donor and the acceptor?

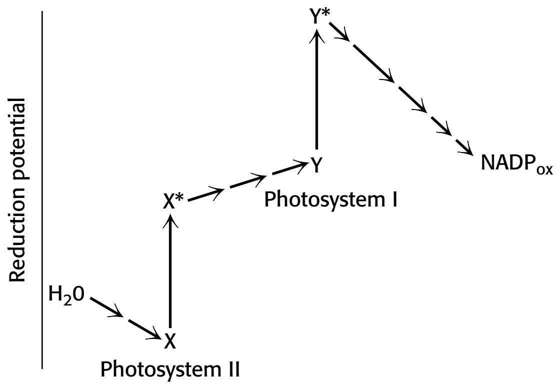
10. *Neutralization compensation*. In chloroplasts, a greater pH gradient across the thylakoid membrane is required to power the synthesis of ATP than is required across the mitochondrial inner membrane. Explain this difference.

11. *Environmentally appropriate*. Chlorophyll is a hydrophobic molecule. Why is this property crucial for the function of chlorophyll?

12. *Proton origins*. What are the various sources of protons that contribute to the generation of a proton gradient in chloroplasts?

13. *Efficiency matters.* What fraction of the energy of 700-nm light absorbed by photosystem I is trapped as high-energy electrons?

14. *That's not right*. Explain the defect or defects in the hypothetical scheme for the light reactions of photosynthesis depicted here.



15. *Electron transfer*. Calculate the $\Delta E'_0$ and $\Delta G^{\circ\prime}$ for the reduction of $NADP^+$ by ferredoxin. Use data given in Table 18.1.

16. *To boldly go*. (a) It can be argued that, if life were to exist elsewhere in the universe, it would require some process like photosynthesis. Why is this argument reasonable? (b) If the starship *Enterprise* were to land on a distant planet and find no measurable oxygen in the atmosphere, could the crew conclude that photosynthesis is not taking place?

17. *Weed killer 1*. Dichlorophenyldimethylurea (DCMU), a herbicide, interferes with photophosphorylation and O_2 evolution. However, it does not block O_2 evolution in the presence of an artificial electron acceptor such as ferricyanide. Propose a site for the inhibitory action of DCMU.

18. *Weed killer 2*. Predict the effect of the herbicide dichlorophenyldimethylurea (DCMU) on a plant's ability to perform cyclic photophosphorylation.

19. *Infrared harvest*. Consider the relation between the energy of a photon and its wavelength.

(a) Some bacteria are able to harvest 1000-nm light. What is the energy (in kilojoules or kilocalories) of a mole (also called an *einstein*) of 1000-nm photons?

(b) What is the maximum increase in redox potential that can be induced by a 1000-nm photon?

(c) What is the minimum number of 1000-nm photons needed to form ATP from ADP and P_i? Assume a ΔG of 50 kJ mol^{-1} (12 kcal mol^{-1}) for the phosphorylation reaction.

20. *Missing acceptors.* Suppose that a bacterial reaction center containing only the special pair and the quinones has been prepared. Given the separation of 22 Å between the special pair and the closest quinone, estimate the rate of electron transfer between the excited special pair and this quinone.

21. *Close approach.* Suppose that energy transfer between two chlorophyll *a* molecules separated by 10 Å takes place in 10 picoseconds. Suppose that this distance is increased to 20 Å with all other factors remaining the same. How long would energy transfer take?

Chapter Integration Problems

22. *Functional equivalents.* What structural feature of mitochondria corresponds to the thylakoid membranes?

23. *Compare and contrast.* Compare and contrast oxidative phosphorylation and photosynthesis.

24. *Energy accounts.* On page 597, the balance sheet for the cost of the photosynthetically powered synthesis of glucose is presented. Eighteen molecules of ATP are required. Yet, when glucose undergoes combustion in cellular respiration, 30 molecules of ATP are produced. Account for the difference.

25. *Looking for a place to rest.* Albert Szent-Györgyi, Nobel Prize–winning biochemist, once said, "Life is nothing but an electron looking for a place to rest." Explain how this pithy statement applies to photosynthesis and cellular respiration.

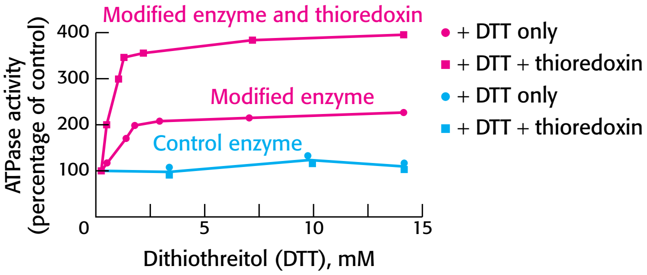
Mechanism Problem

26. *Hill reaction.* In 1939, Robert Hill discovered that chloroplasts evolve O_2 when they are illuminated in the presence of an artificial electron acceptor such as ferricyanide $[Fe^{3+}(CN)_6]^{3-}$. Ferricyanide is reduced to ferrocyanide $[Fe^{2+}(CN)_6]^{4-}$ in this process. No NADPH or reduced plastocyanin is produced. Propose a mechanism for the Hill reaction.

Data Interpretation and Chapter Integration Problems

27. *The same, but different.* The $\alpha_3\beta_3\gamma$ complex of mitochondrial or chloroplast ATP synthase functions as an ATPase in vitro. The chloroplast enzyme (both synthase and ATPase activity) is sensitive to redox control, whereas the mitochondrial enzyme is not. To determine where the enzymes differ, a segment of the mitochondrial γ subunit was removed and replaced with the equivalent segment from the chloroplast γ subunit. The ATPase activity of the modified enzyme was then measured as a function of redox conditions. Dithiothreitol (DTT) is a reducing agent.

(a) What is the redox regulator of the chloroplast ATP synthase in vivo? The adjoining graph shows the ATPase activity of modified and control enzymes under various redox conditions.



[Data from O. Bald et al. *J. Biol. Chem.* 275:12757–12762, 2000.]

(b) What is the effect of increasing the reducing power of the reaction mixture for the control and the modified enzymes?

(c) What is the effect of the addition of thioredoxin? How do these results differ from those in the presence of DTT alone? Suggest a possible explanation for the difference.

(d) Did the researchers succeed in identifying the region of the γ subunit responsible for redox regulation?

(e) What is the biological rationale of regulation by high concentrations of reducing agents?

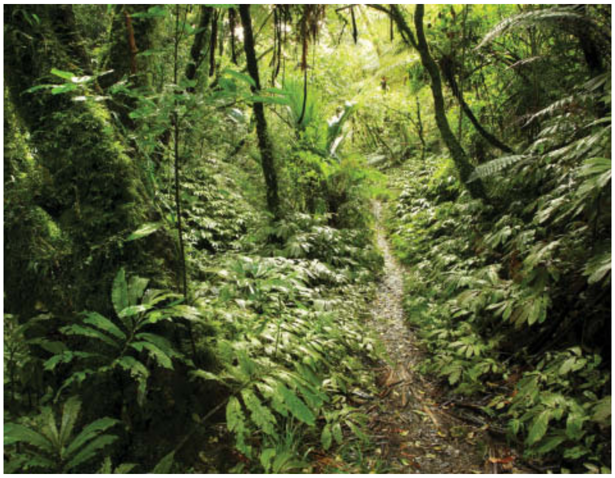
(f) What amino acids in the γ subunit are most likely affected by the reducing conditions?

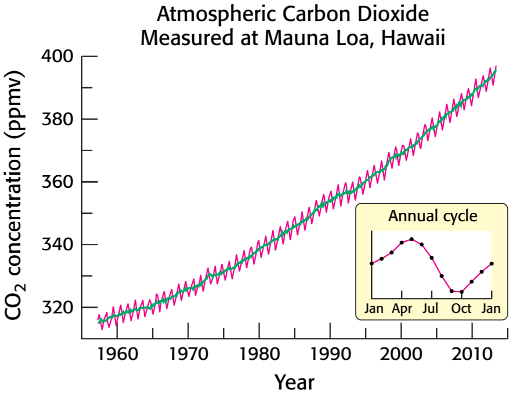
(g) What experiments might confirm your answer to part *e*?

CHAPTER

20

The Calvin Cycle and the Pentose Phosphate Pathway




Atmospheric Carbon Dioxide
Measured at Mauna Loa, Hawaii
CO_2 concentration (ppmv)
400
380
360
340
320
1960
1970
1980
1990
2000
2010
Year
Annual cycle
Jan
Apr
Jul
Oct
Jan

Atmospheric carbon dioxide measurements at Mauna Loa, Hawaii. These measurements show the increase in carbon dioxide since 1960. The sawtooth appearance of the curve is the result of annual cycles resulting from seasonal variation in CO_2 fixation by the Calvin cycle in terrestrial plants (see insert). Much of this fixation takes place in rain forests, which account for approximately 50% of terrestrial fixation. [(Left) stillfx/ age fotostock. (Right) Data from http://www.esrl.noaa.gov/gmd/ccgg/trends]

Photosynthesis proceeds in two parts: the light reactions and the dark reactions. The light reactions, discussed in Chapter 19, transform light energy into ATP and biosynthetic reducing power, NADPH. The dark reactions use the ATP and NADPH produced by the light reactions to reduce carbon atoms from their fully oxidized state as carbon dioxide to a more reduced state as a hexose. Carbon dioxide is thereby trapped in a form that is useful for many processes and most especially as a fuel. Together, *the light reactions and dark reactions of photosynthesis cooperate to transform light energy into carbon fuel*. The dark reactions are called either the *Calvin–Benson cycle,* after Melvin Calvin and Andrew Benson, the biochemists who elucidated the pathway, or simply the Calvin cycle. The components of the Calvin cycle are called the *dark reactions* because, in contrast with the light reactions, these reactions do not directly depend on the presence of light.

The second half of this chapter examines a pathway common to all organisms, known variously as the *pentose phosphate pathway,* the *hexose monophosphate pathway,* the *phosphogluconate pathway,* or the *pentose shunt.* The pathway provides a means by which glucose can be oxidized to generate NADPH, *the currency of readily available reducing power in cells*. The phosphoryl group on the 2′-hydroxyl group of one of the ribose units of NADPH distinguishes NADPH from NADH. *There is a fundamental distinction*

OUTLINE

20.1 The Calvin Cycle Synthesizes Hexoses from Carbon Dioxide and Water

20.2 The Activity of the Calvin Cycle Depends on Environmental Conditions

20.3 The Pentose Phosphate Pathway Generates NADPH and Synthesizes Five-Carbon Sugars

20.4 The Metabolism of Glucose 6-phosphate by the Pentose Phosphate Pathway Is Coordinated with Glycolysis

20.5 Glucose 6-phosphate Dehydrogenase Plays a Key Role in Protection Against Reactive Oxygen Species

between NADPH and NADH in biochemistry: NADH is oxidized by the respiratory chain to generate ATP, whereas NADPH serves as a reductant in biosynthetic processes. The pentose phosphate pathway can also be used for the catabolism of pentose sugars from the diet, the synthesis of pentose sugars for nucleotide biosynthesis, and the catabolism and synthesis of less common four- and seven-carbon sugars. The pentose phosphate pathway and the Calvin cycle have in common several enzymes and intermediates that attest to an evolutionary kinship. Like glycolysis and gluconeogenesis, these pathways are mirror images of each other: the Calvin cycle uses NADPH to reduce carbon dioxide to generate hexoses, whereas the pentose phosphate pathway oxidizes glucose to carbon dioxide to generate NADPH.

20.1 The Calvin Cycle Synthesizes Hexoses from Carbon Dioxide and Water

As stated in Chapter 16, glucose can be formed from noncarbohydrate precursors, such as lactate and amino acids, by gluconeogenesis. The energy powering gluconeogenesis ultimately comes from previous catabolism of carbon fuels. In contrast, photosynthetic organisms can use the Calvin cycle to synthesize glucose from carbon dioxide gas and water, by using sunlight as an energy source. The Calvin cycle introduces into life all the carbon atoms that will be used as fuel and as the carbon backbones of biomolecules. Photosynthetic organisms are called *autotrophs* (literally, *self-feeders*) because they can convert sunlight into chemical energy, which they subsequently use to power their biosynthetic processes. Organisms that obtain energy from chemical fuels only are called *heterotrophs,* and such organisms ultimately depend on autotrophs for their fuel.

The Calvin cycle comprises three stages (Figure 20.1):

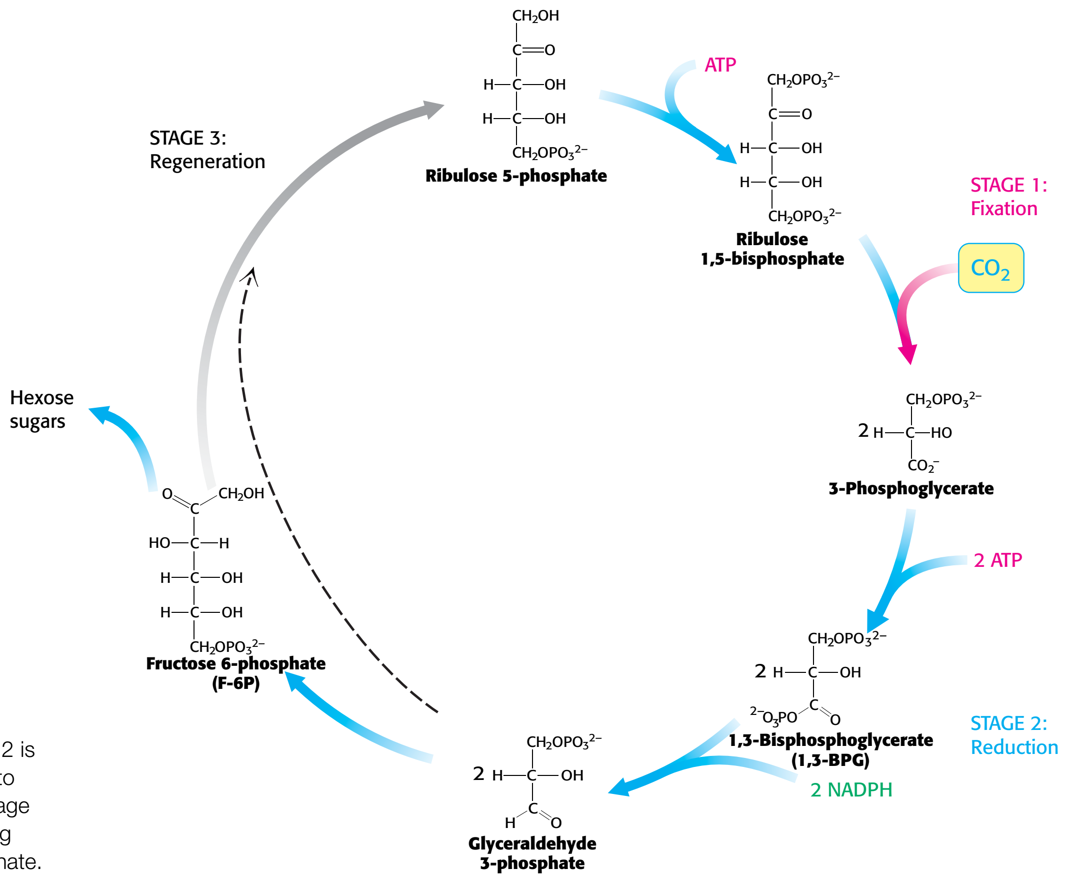

Ribulose 5-phosphate
ATP
Ribulose 1,5-bisphosphate
STAGE 1: Fixation
CO$_2$
2
3-Phosphoglycerate
2 ATP
2
1,3-Bisphosphoglycerate (1,3-BPG)
STAGE 2: Reduction
2 NADPH
2
Glyceraldehyde 3-phosphate
Fructose 6-phosphate (F-6P)
Hexose sugars
STAGE 3: Regeneration

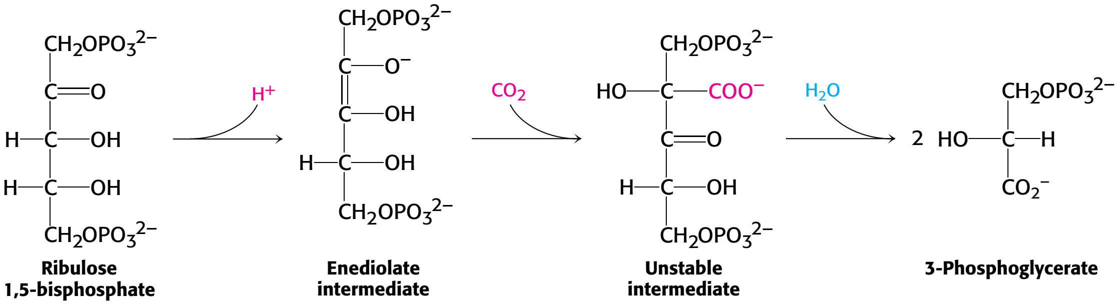
FIGURE 20.1 Calvin cycle. The Calvin cycle consists of three stages. Stage 1 is the fixation of carbon by the carboxylation of ribulose 1,5-bisphosphate. Stage 2 is the reduction of the fixed carbon to begin the synthesis of hexose. Stage 3 is the regeneration of the starting compound ribulose 1,5-bisphosphate.

1. The fixation of CO_2 by ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate;
2. The reduction of 3-phosphoglycerate to form hexose sugars; and
3. The regeneration of ribulose 1,5-bisphosphate so that more CO_2 can be fixed.

This set of reactions takes place in the stroma of chloroplasts, the photosynthetic organelles.

Carbon dioxide reacts with ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate

The first step in the Calvin cycle is the fixation of CO_2. This begins with the conversion of ribulose 1,5-bisphosphate into a highly reactive enediolate intermediate. The CO_2 molecule condenses with the enediol intermediate to form an unstable six-carbon compound, which is rapidly hydrolyzed to two molecules of 3-phosphoglycerate.


$CH_2OPO_3^{2-}$
C=O
H—C—OH
H—C—OH
$CH_2OPO_3^{2-}$
Ribulose 1,5-bisphosphate
H^+
$CH_2OPO_3^{2-}$
C—O^-
C—OH
H—C—OH
$CH_2OPO_3^{2-}$
Enediolate intermediate
CO_2
$CH_2OPO_3^{2-}$
HO—C—COO^-
C=O
H—C—OH
$CH_2OPO_3^{2-}$
Unstable intermediate
H_2O
2 HO—C—H
$CH_2OPO_3^{2-}$
CO_2^-
3-Phosphoglycerate

This highly exergonic reaction [$\Delta G^{\circ\prime} = -51.9\ \text{kJ mol}^{-1}\ (-12.4\ \text{kcal mol}^{-1})$] is catalyzed by *ribulose 1,5-bisphosphate carboxylase/oxygenase* (usually called *rubisco*), an enzyme located on the stromal surface of the thylakoid membranes of chloroplasts. This important reaction is the rate-limiting step in hexose synthesis. In plants and green algae, rubisco consists of eight large (L, 55-kDa) subunits and eight small (S, 15-kDa) ones (Figure 20.2). Each L chain contains a catalytic site and a regulatory site. The S chains enhance the catalytic activity of the L chains. This enzyme is abundant in chloroplasts, accounting for approximately 30% of the total leaf protein in some plants. In fact, rubisco is the most abundant enzyme and probably the most abundant protein in the biosphere. Large amounts are present because rubisco is a slow enzyme; its maximal catalytic rate is only 3 s^{-1}.

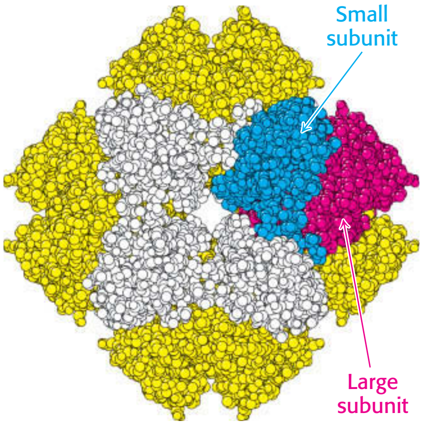
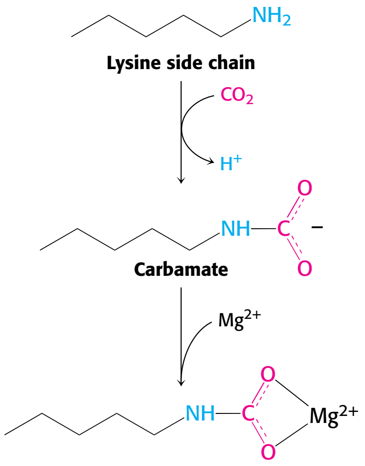

Small subunit
Large subunit

FIGURE 20.2 Structure of rubisco. The enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco) comprises eight large subunits (one shown in red and the others in yellow) and eight small subunits (one shown in blue and the others in white). The active sites lie in the large subunits. [Drawn from 1RXO.pdb.]


NH₂
Lysine side chain
CO₂
H⁺
NH—C
O
O
–
Carbamate
Mg²⁺
NH—C
O
O
Mg²⁺

Rubisco activity depends on magnesium and carbamate

Rubisco requires a bound divalent metal ion for activity, usually magnesium ion. Like the zinc ion in the active site of carbonic anhydrase (Section 9.2), this metal ion serves to activate a bound substrate molecule by stabilizing a negative charge. Interestingly, a CO_2 molecule other than the substrate is required to complete the assembly of the Mg^{2+}-binding site in rubisco. This CO_2 molecule adds to the uncharged ε-amino group of lysine 201 to form a *carbamate*. This negatively charged adduct then binds the Mg^{2+} ion.

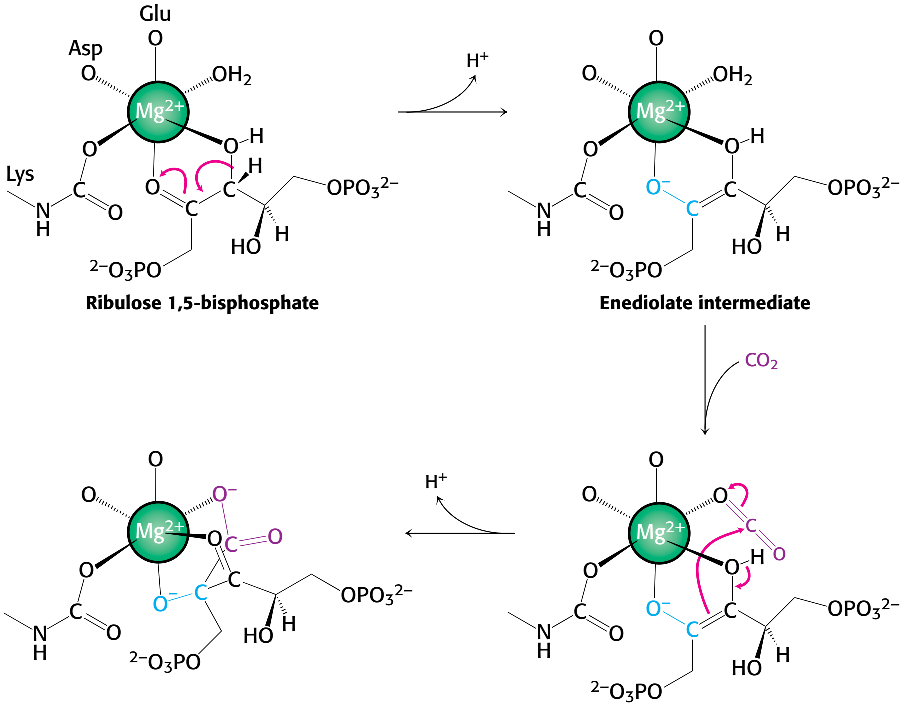
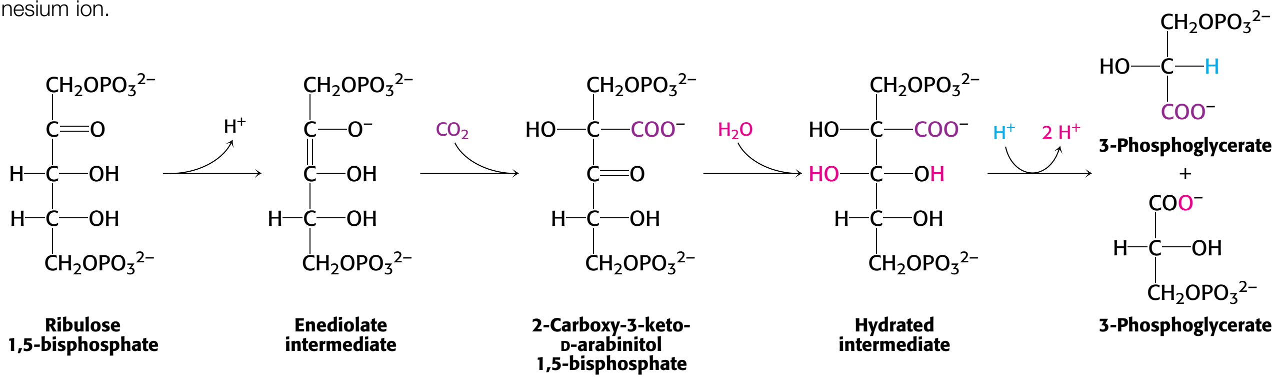

Glu
Asp
Lys
Mg²⁺
OH₂
OPO₃²⁻
²⁻O₃PO
HO
H
Ribulose 1,5-bisphosphate
H⁺
Enediolate intermediate
CO₂
H⁺

FIGURE 20.3 Role of the magnesium ion in the rubisco mechanism. Ribulose 1,5-bisphosphate binds to a magnesium ion that is linked to rubisco through a glutamate residue, an aspartate residue, and the lysine carbamate. The coordinated ribulose 1,5-bisphosphate gives up a proton to form a reactive enediolate species that reacts with CO_2 to form a new carbon–carbon bond.

The metal center plays a key role in binding ribulose 1,5-bisphosphate and activating it so that it reacts with CO_2 (Figure 20.3). Ribulose 1,5-bisphosphate binds to Mg^{2+} through its keto group and an adjacent hydroxyl group. This complex is readily deprotonated to form an enediolate intermediate. This reactive species, analogous to the zinc–hydroxide species in carbonic anhydrase, couples with CO_2, forming the new carbon–carbon bond. The resulting 2-carboxy-3-keto-D-arabinitol 1,5-bisphosphate is coordinated to the Mg^{2+} ion through three groups, including the newly formed carboxylate. A molecule of H_2O is then added to this β-ketoacid to form an intermediate that cleaves to form two molecules of 3-phosphoglycerate (Figure 20.4).

FIGURE 20.4 Formation of 3-phosphoglycerate. The overall pathway for the conversion of ribulose 1,5 bisphosphate and CO_2 into two molecules of 3-phosphoglycerate. Although the free species are shown, these steps take place on the magnesium ion.


Ribulose 1,5-bisphosphate
H⁺
Enediolate intermediate
CO₂
2-Carboxy-3-keto-D-arabinitol 1,5-bisphosphate
H₂O
Hydrated intermediate
H⁺
2 H⁺
3-Phosphoglycerate
+
3-Phosphoglycerate

Rubisco activase is essential for rubisco activity

In the absence of CO_2, rubisco binds to its substrate, ribulose 1,5-bisphosphate, so tightly that its enzyme activity is inhibited. The enzyme *rubisco activase* uses ATP to induce structural changes in rubisco that allow release of the bound substrate and formation of the required carbamate. The ATP requirement of rubisco activase coordinates rubisco activity with the light reactions.

The activase contains a P-loop domain (Section 9.4) and assembles into a hexameric ring–like structure. The enzyme binds to the C-terminal end of the polypeptide chain that forms that large subunit. Using the energy of ATP hydrolysis, the C-terminal is transiently pulled in to the central pore of the activase. This releases the bound ribulose 1,5-bisphosphate and activates rubisco. Activase is a member of the large AAA family of ATPases that usually assemble into ring-shaped oligomeric structures and use the ATP hydrolysis to power conformational changes in their substrates. AAA ATPases play a variety of roles in biological systems. They are crucial for protein degradation (Chapter 23) and DNA replication (Chapter 28), and also function as molecular motors (Chapter 35).

Rubisco also catalyzes a wasteful oxygenase reaction: Catalytic imperfection

The reactive enediolate intermediate generated on the Mg^{2+} ion sometimes reacts with O_2 instead of CO_2; thus, rubisco also catalyzes a deleterious oxygenase reaction. The products of this reaction are *phosphoglycolate* and *3-phosphoglycerate* (Figure 20.5). The rate of the carboxylase reaction is four times that of the oxygenase reaction under normal atmospheric conditions at 25°C; the stromal concentration of CO_2 is then 10 μM and that of O_2 is 250 μM. The oxygenase reaction, like the carboxylase reaction, requires that lysine 201 be in the carbamate form. Because this carbamate forms only in the presence of CO_2, rubisco is prevented from catalyzing the oxygenase reaction exclusively when CO_2 is absent.

Phosphoglycolate is not a versatile metabolite. A salvage pathway recovers part of its carbon skeleton (Figure 20.6). A specific phosphatase converts phosphoglycolate into *glycolate,* which enters *peroxisomes* (also called *microbodies;* Figure 20.7). Glycolate is then oxidized to *glyoxylate* by glycolate oxidase, an enzyme with a flavin mononucleotide prosthetic group. The H_2O_2 produced in this reaction is cleaved by catalase to H_2O and O_2. Transamination of glyoxylate then yields *glycine,* which then enters the mitochondria. Two glycine molecules can unite to form serine with the release of CO_2 and ammonium ion (NH_4^+). The ammonium ion, used in the synthesis of nitrogen-containing compounds, is salvaged by a glutamine synthetase reaction (Figure 20.6 and Section 23.3). Serine reenters the peroxisome,where it donates its ammonia to glyoxylate and is subsequently converted back into 3-phosphoglycerate.

This salvage pathway serves to recycle three of the four carbon atoms of two molecules of glycolate. However, one carbon atom is lost as CO_2. This

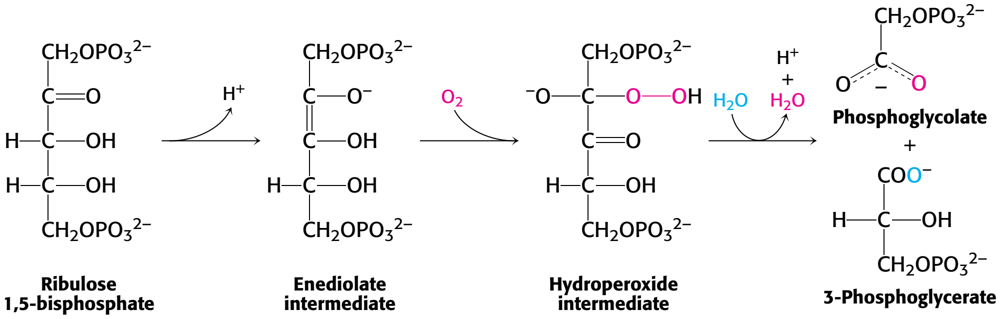


FIGURE 20.5 A wasteful side reaction. The reactive enediolate intermediate on rubisco also reacts with molecular oxygen to form a hydroperoxide intermediate, which then proceeds to form one molecule of 3-phosphoglycerate and one molecule of phosphoglycolate.

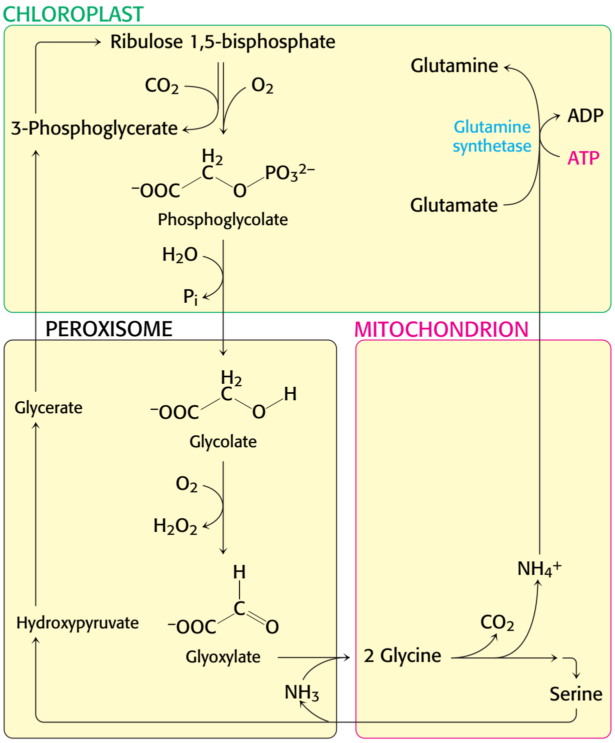

CHLOROPLAST
Ribulose 1,5-bisphosphate
CO2
O2
3-Phosphoglycerate
H2
C
-OOC
O
PO3 2−
Phosphoglycolate
H2O
Pi
Glutamine
Glutamine synthetase
ADP
ATP
Glutamate
PEROXISOME
Glycerate
H2
C
-OOC
O
H
Glycolate
O2
H2O2
H
C
-OOC
O
Glyoxylate
Hydroxypyruvate
NH3
MITOCHONDRION
NH4+
CO2
2 Glycine
Serine

FIGURE 20.6 Photorespiratory reactions. Phosphoglycolate is formed as a product of the oxygenase reaction in chloroplasts. After dephosphorylation, glycolate is transported into peroxisomes where it is converted into glyoxylate and then glycine. In mitochondria, two glycines are converted into serine, after losing a carbon as CO_2 and ammonium ion. The serine is converted back into 3-phosphoglycerate and the ammonium ion is salvaged in chloroplasts.

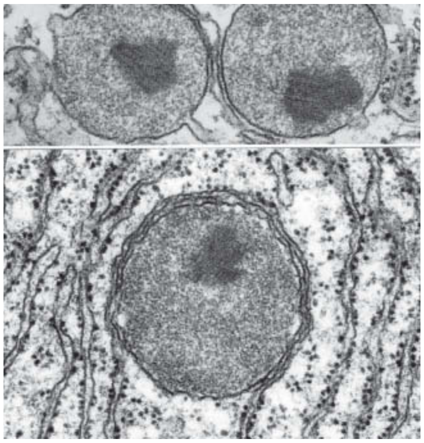


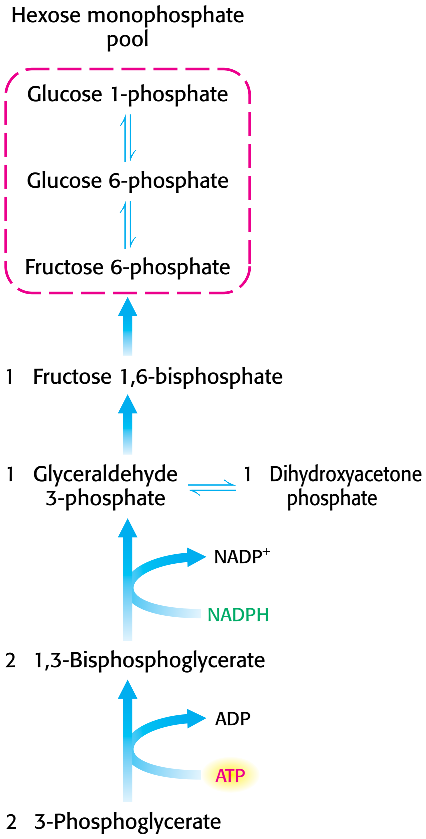
FIGURE 20.7 Electron micrograph of three peroxisomes. The peroxisomes are surrounded by rough endoplasmic reticulum. [Don W. Fawcett/Science Source.]

process is called *photorespiration* because O_2 is consumed and CO_2 is released. Photorespiration is wasteful because organic carbon is converted into CO_2 without the production of ATP, NADPH, or another energy-rich metabolite. Although evolutionary processes have presumably enhanced the preference of rubisco for carboxylation—for instance, the rubisco of higher plants is eightfold as specific for carboxylation as that of photosynthetic bacteria—photorespiration still accounts for the loss of up to 25% of the carbon fixed.

Much research has been undertaken to generate recombinant forms of rubisco that display reduced oxygenase activity, but all such attempts have failed. This raises the following question: What is the biochemical basis of this inefficiency? Structural studies show that when the reactive enediolate intermediate is formed, loops close over the active site to protect the enediolate. A channel to the environment is maintained to allow access by CO_2. However, like CO_2, O_2 is a linear molecule that also fits the channel. In essence, the problem lies not with the enzyme but in the unremarkable structure of CO_2. CO_2 lacks any chemical features that would allow discrimination between it and other gases such as O_2, and thus the oxygenase activity is an inevitable failing of the enzyme. Another possibility exists, however. The oxygenase activity may not be an imperfection of the enzyme, but rather an imperfection in our understanding. Perhaps the oxygenase activity performs a biochemically important role that we have not yet discovered.

Hexose phosphates are made from phosphoglycerate, and ribulose 1,5-bisphosphate is regenerated

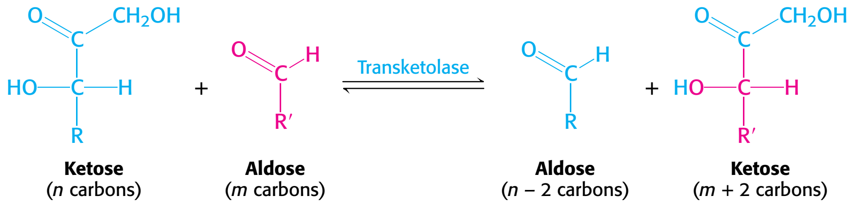
The 3-phosphoglycerate product of rubisco is next converted into fructose 6-phosphate, which readily isomerizes to glucose 1-phosphate and glucose 6-phosphate. The mixture of the three phosphorylated hexoses is called the *hexose monophosphate pool*. The steps in this conversion (Figure 20.8) are like those of the gluconeogenic pathway (Figure 16.24), except that glyceraldehyde 3-phosphate dehydrogenase in chloroplasts, which generates glyceraldehyde 3-phosphate (GAP), is specific for NADPH rather than NADH. These reactions and that catalyzed by rubisco bring CO_2 to the level of a hexose, converting CO_2 into a chemical fuel at the expense of NADPH and ATP generated from the light reactions.



Hexose monophosphate pool
Glucose 1-phosphate
Glucose 6-phosphate
Fructose 6-phosphate
1 Fructose 1,6-bisphosphate
1 Glyceraldehyde 3-phosphate
1 Dihydroxyacetone phosphate
NADP+
NADPH
2 1,3-Bisphosphoglycerate
ADP
ATP
2 3-Phosphoglycerate

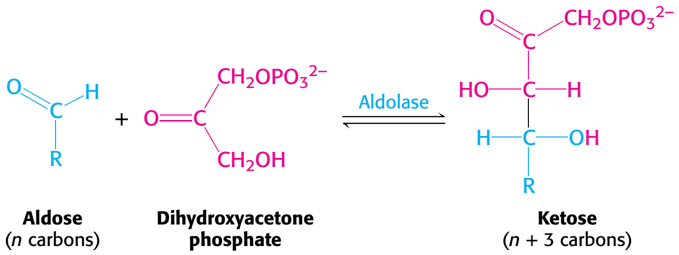
FIGURE 20.8 Hexose phosphate formation. 3-Phosphoglycerate is converted into fructose 6-phosphate in a pathway parallel to that of gluconeogenesis.

The third phase of the Calvin cycle is the regeneration of ribulose 1,5-bisphosphate, the acceptor of CO_2 in the first step. The problem is to construct a five-carbon sugar from six-carbon and three-carbon sugars. A transketolase and an aldolase play the major role in the rearrangement of the carbon atoms. The *transketolase,* which we will see again in the pentose phosphate pathway, requires the coenzyme thiamine pyrophosphate (TPP) to transfer a two-carbon unit ($CO—CH_2OH$) from a ketose to an aldose.



Transketolase
Ketose
(n carbons)
Aldose
(m carbons)
Aldose
(n − 2 carbons)
Ketose
(m + 2 carbons)

Aldolase, which we have already encountered in glycolysis (Section 16.1), catalyzes an aldol condensation between dihydroxyacetone phosphate (DHAP) and an aldehyde. This enzyme is highly specific for dihydroxyacetone phosphate, but it accepts a wide variety of aldehydes.



Aldolase
Aldose
(n carbons)
Dihydroxyacetone phosphate
Ketose
(n + 3 carbons)

With these enzymes, the construction of the five-carbon sugar proceeds as shown in Figure 20.9.

Finally, ribose 5-phosphate is converted into ribulose 5-phosphate by *phosphopentose isomerase,* whereas xylulose 5-phosphate is converted into ribulose 5-phosphate by *phosphopentose epimerase*. Ribulose 5-phosphate is converted into ribulose 1,5-bisphosphate through the action of *phosphoribulokinase* (Figure 20.10). The sum of these reactions is

$$\begin{aligned}&\text{Fructose 6-phosphate} + 2\text{ glyceraldehyde 3-phosphate}\\&\qquad + \text{dihydroxyacetone phosphate} + 3\text{ ATP} \longrightarrow\\&\qquad\qquad 3\text{ ribulose 1,5-bisphosphate} + 3\text{ ADP}\end{aligned}$$

Figure 20.11 presents the required reactions with the proper stoichiometry to convert three molecules of CO_2 into one molecule of DHAP. However, two molecules of DHAP are required for the synthesis of a member of the hexose monophosphate pool. Consequently, the cycle as presented must take place twice to yield a hexose monophosphate. The outcome of the Calvin cycle is the generation of a hexose and the regeneration of the starting compound, ribulose 1,5-bisphosphate. In essence, ribulose 1,5-bisphosphate acts catalytically, similarly to oxaloacetate in the citric acid cycle.

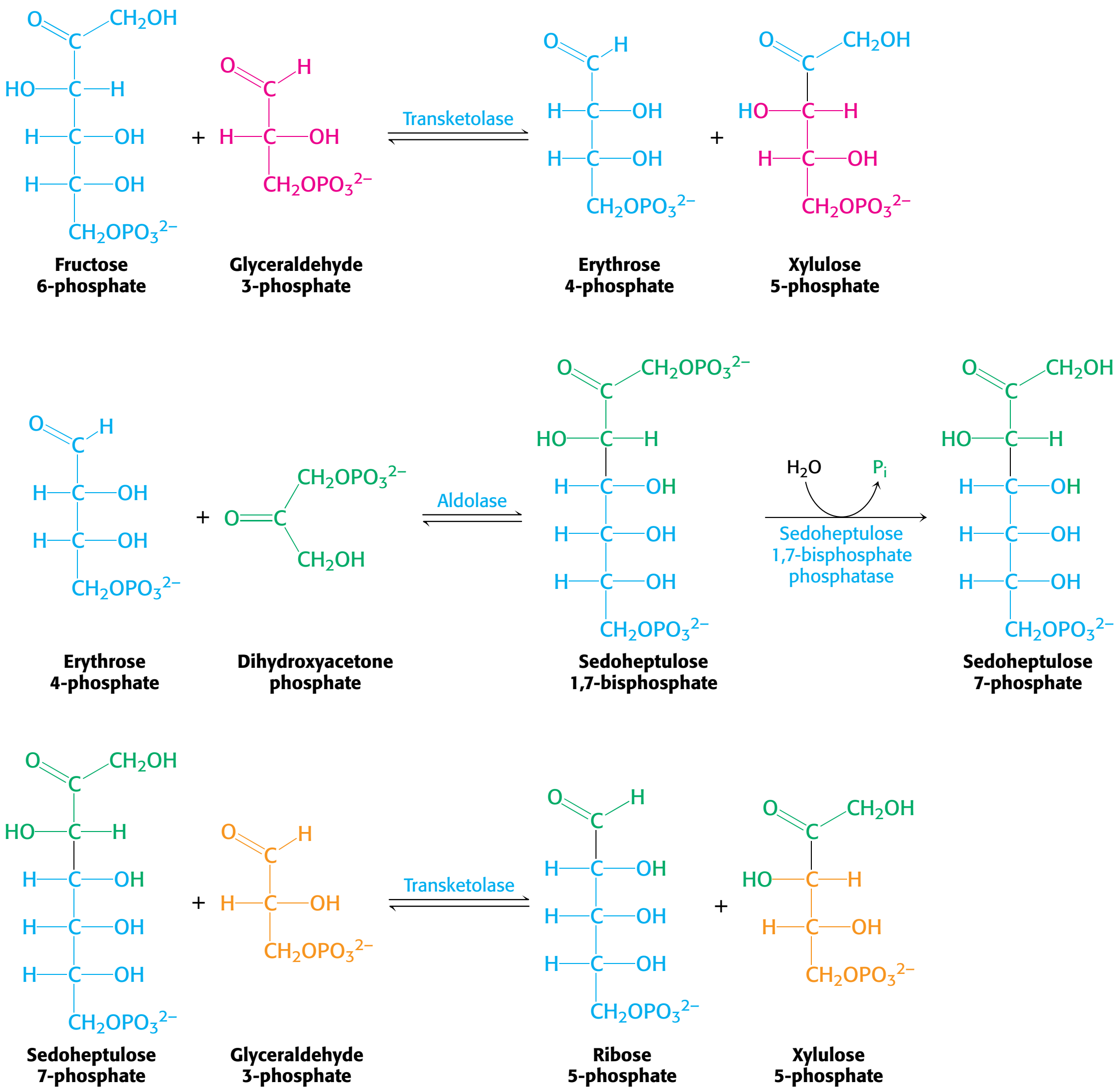

Fructose 6-phosphate
Glyceraldehyde 3-phosphate
Transketolase
Erythrose 4-phosphate
Xylulose 5-phosphate
Erythrose 4-phosphate
Dihydroxyacetone phosphate
Aldolase
Sedoheptulose 1,7-bisphosphate
H_2O
P_i
Sedoheptulose 1,7-bisphosphate phosphatase
Sedoheptulose 7-phosphate
Sedoheptulose 7-phosphate
Glyceraldehyde 3-phosphate
Transketolase
Ribose 5-phosphate
Xylulose 5-phosphate

FIGURE 20.9 Formation of five-carbon sugars. First, transketolase converts a six-carbon sugar and a three-carbon sugar into a four-carbon sugar and a five-carbon sugar. Then, aldolase combines the four-carbon product and a three-carbon sugar to form a seven-carbon sugar. Finally, this seven-carbon sugar reacts with another three-carbon sugar to form two additional five-carbon sugars.

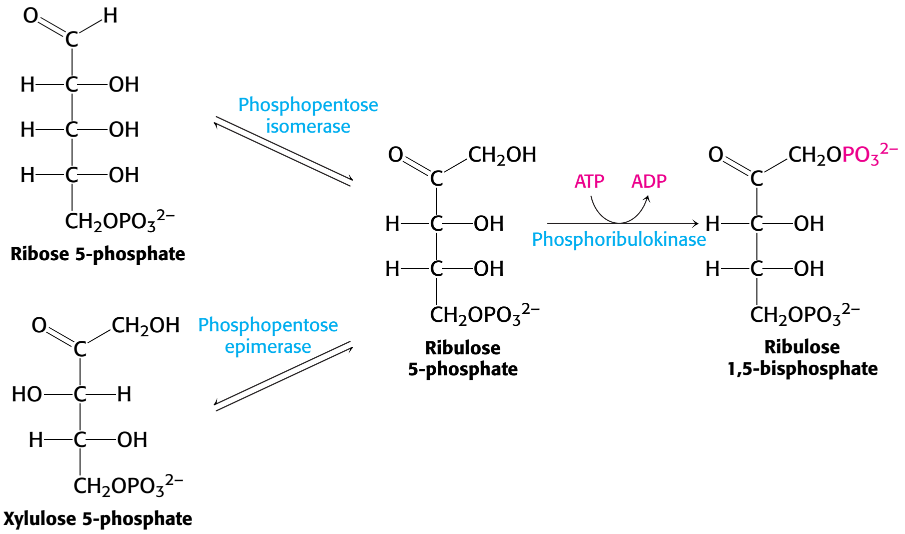

Ribose 5-phosphate
Phosphopentose isomerase
Xylulose 5-phosphate
Phosphopentose epimerase
Ribulose 5-phosphate
ATP
ADP
Phosphoribulokinase
Ribulose 1,5-bisphosphate

FIGURE 20.10 Regeneration of ribulose 1,5-bisphosphate. Both ribose 5-phosphate and xylulose 5-phosphate are converted into ribulose 5-phosphate, which is then phosphorylated to complete the regeneration of ribulose 1,5-bisphosphate.

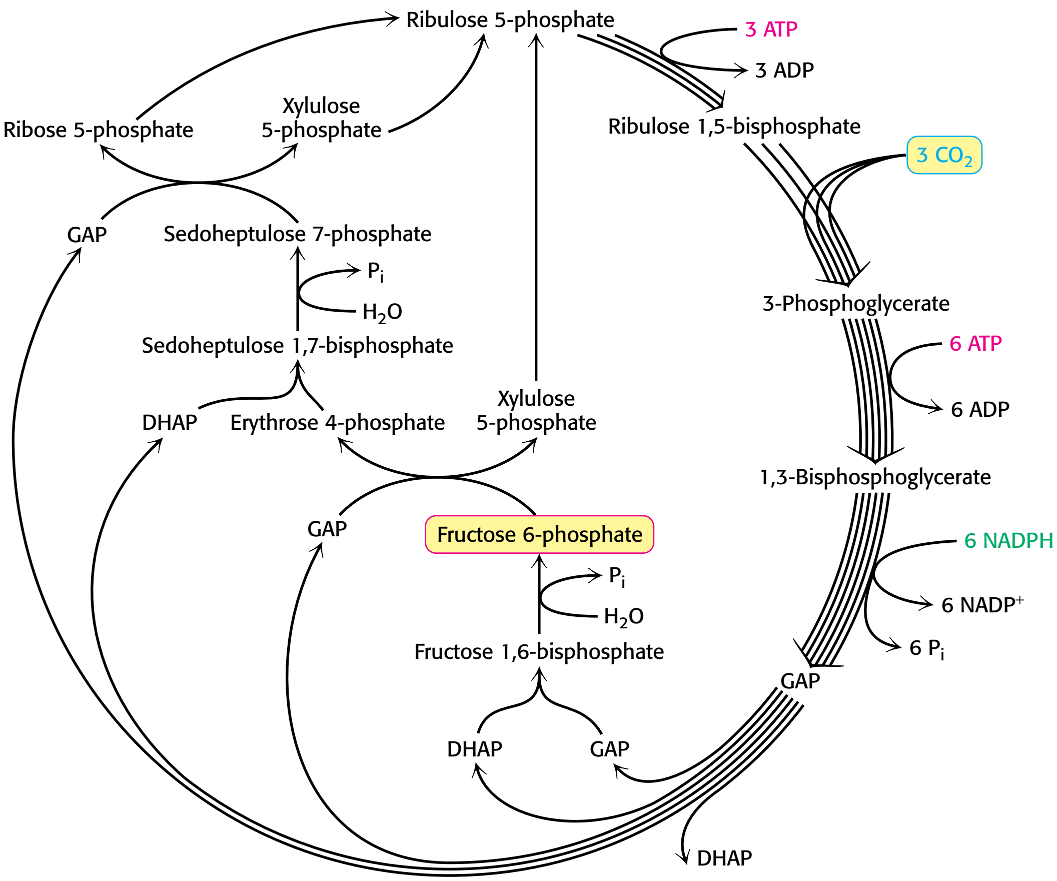

Ribulose 5-phosphate
3 ATP
3 ADP
Ribose 5-phosphate
Xylulose 5-phosphate
Ribulose 1,5-bisphosphate
3 CO$_2$
GAP
Sedoheptulose 7-phosphate
P$_i$
H$_2$O
Sedoheptulose 1,7-bisphosphate
3-Phosphoglycerate
6 ATP
6 ADP
DHAP
Erythrose 4-phosphate
Xylulose 5-phosphate
1,3-Bisphosphoglycerate
GAP
Fructose 6-phosphate
6 NADPH
P$_i$
H$_2$O
6 NADP$^+$
6 P$_i$
Fructose 1,6-bisphosphate
GAP
DHAP
GAP
DHAP

FIGURE 20.11 Calvin cycle. The diagram shows the reactions necessary with the correct stoichiometry to convert three molecules of CO$_2$ into one molecule of dihydroxyacetone phosphate (DHAP). The cycle is not as simple as presented in Figure 20.1; rather, it entails many reactions that lead ultimately to the synthesis of glucose and the regeneration of ribulose 1,5-bisphosphate. [Information from J. R. Bowyer and R. C. Leegood. "Photosynthesis," in *Plant Biochemistry,* P. M. Dey and J. B. Harborne, Eds. (Academic Press, 1997), p. 85.]

Three ATP and two NADPH molecules are used to bring carbon dioxide to the level of a hexose

What is the energy expenditure for synthesizing a hexose? Six rounds of the Calvin cycle are required, because one carbon atom is reduced in each round. Twelve molecules of ATP are expended in phosphorylating 12 molecules of 3-phosphoglycerate to 1,3-bisphosphoglycerate, and 12 molecules of NADPH are consumed in reducing 12 molecules of 1,3-bisphosphoglycerate to glyceraldehyde 3-phosphate. An additional six molecules of ATP are spent in regenerating ribulose 1,5-bisphosphate.

We can now write a balanced equation for the net reaction of the Calvin cycle:

$$\begin{aligned}6\,CO_2 + 18\,ATP + 12\,NADPH + 12\,H_2O \longrightarrow \\ C_6H_{12}O_6 + 18\,ADP + 18\,P_i + 12\,NADP^+ + 6\,H^+\end{aligned}$$

Thus, three molecules of ATP and two molecules of NADPH are consumed in incorporating a single CO$_2$ molecule into a hexose such as glucose or fructose.

Starch and sucrose are the major carbohydrate stores in plants

What are the fates of the members of the hexose monophosphate pool? These molecules are used in a variety of ways, but there are two primary roles. Plants contain two major storage forms of sugar: *starch* and *sucrose*. Starch, like its animal counterpart glycogen, is a polymer of glucose residues, but it is less branched than glycogen because it contains a smaller proportion of α-1,6-glycosidic linkages (Section 11.2). Another difference is that ADP-glucose, not UDP-glucose, is the activated precursor. Starch is synthesized and stored in chloroplasts.

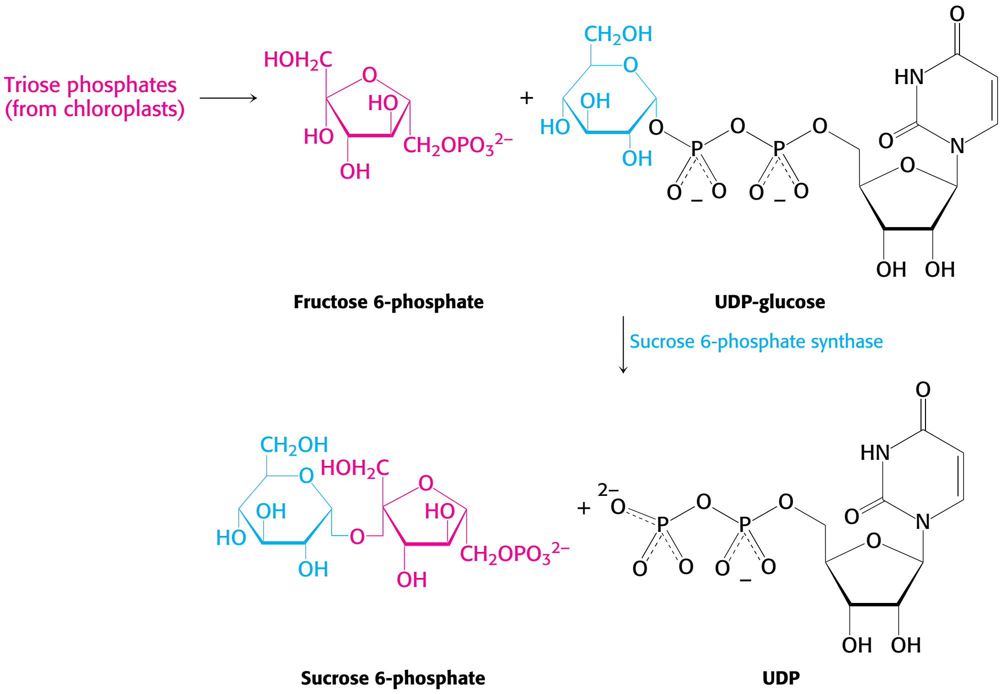


FIGURE 20.12 Synthesis of sucrose. Sucrose 6-phosphate is formed by the reaction between fructose 6-phosphate and the activated intermediate uridine diphosphate glucose (UDP-glucose). Sucrose phosphatase subsequently generates free sucrose (not shown).

In contrast, sucrose (common table sugar), a disaccharide, is synthesized in the cytoplasm. Plants lack the ability to transport hexose phosphates across the chloroplast membrane, but they are able to transport triose phosphates from chloroplasts to the cytoplasm. Triose phosphate intermediates such as glyceraldehyde 3-phosphate cross into the cytoplasm in exchange for phosphate through the action of an abundant triose phosphate-phosphate antiporter. Fructose 6-phosphate formed from triose phosphates joins the glucose unit of UDP-glucose to form sucrose 6-phosphate (Figure 20.12). The hydrolysis of the phosphate ester by sucrose phosphatase yields sucrose, a readily transportable and mobilizable sugar that is stored in many plant cells, such as those in sugar beets and sugar cane.

20.2 The Activity of the Calvin Cycle Depends on Environmental Conditions

How do the light reactions communicate with the dark reactions to regulate this crucial process of fixing CO_2 into biomolecules? *The principal means of regulation is alteration of the stromal environment by the light reactions.* The light reactions lead to an increase in stromal pH (a decrease in the H^+ concentration), as well as an increase in the stromal concentrations of Mg^{2+}, NADPH, and reduced ferredoxin—changes that contribute to the activation of certain Calvin-cycle enzymes, located in the stroma (Figure 20.13).

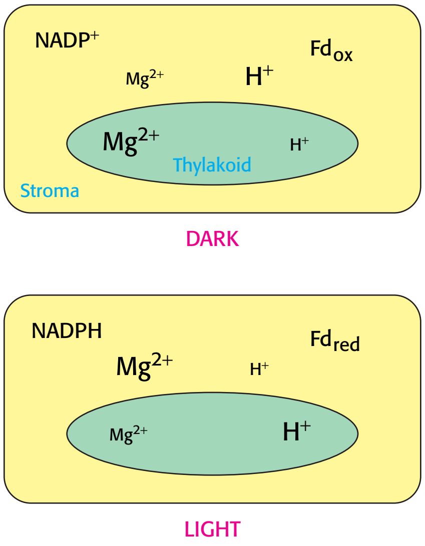


FIGURE 20.13 Light regulation of the Calvin cycle. The light reactions of photosynthesis transfer electrons out of the thylakoid lumen into the stroma and transfer protons from the stroma into the thylakoid lumen. As a consequence of these processes, the concentrations of NADPH, reduced ferredoxin (Fd), and Mg^{2+} in the stroma are higher in the light than in the dark. The stromal pH is also increased (the concentration of H^+ is lowered) in the light as a result of proton pumping from the stroma to the thylakoid lumen. Each of these concentration changes helps couple the Calvin cycle reactions to the light reactions.

Rubisco is activated by light-driven changes in proton and magnesium ion concentrations

As stated earlier, the rate-limiting step in the Calvin cycle is the carboxylation of ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate. *The activity of rubisco increases markedly on illumination because light facilitates the carbamate formation necessary for enzyme activity* (p. 593). In the stroma, the pH increases from 7 to 8, and the level of Mg^{2+}rises. Both effects are consequences of the light-driven pumping of protons into the thylakoid space. Mg^{2+} ions from the thylakoid space are released into the

TABLE 20.1 Enzymes regulated by thioredoxin

Enzyme	Pathway
Rubisco	Carbon fixation in the Calvin cycle
Fructose 1,6-bisphosphatase	Gluconeogenesis
Glyceraldehyde 3-phosphate dehydrogenase	Calvin cycle, gluconeogenesis, glycolysis
Sedoheptulose 1,7-bisphosphatase	Calvin cycle
Glucose 6-phosphate dehydrogenase	Pentose phosphate pathway
Phenylalanine ammonia lyase	Lignin synthesis
Phosphoribulokinase	Calvin cycle
NADP$^+$-malate dehydrogenase	C_4 pathway
CF_1–CF_0 ATP synthase	The light reactions

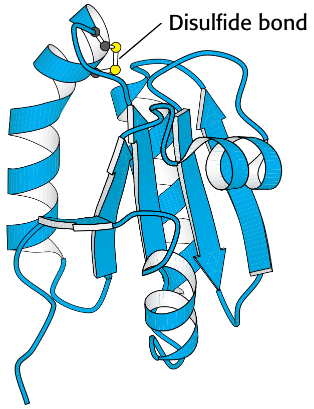


FIGURE 20.14 Thioredoxin. The oxidized form of thioredoxin contains a disulfide bond. When thioredoxin is reduced by reduced ferredoxin, the disulfide bond is converted into two free sulfhydryl groups. Reduced thioredoxin can cleave disulfide bonds in enzymes, activating certain Calvin cycle enzymes and inactivating some degradative enzymes. [Drawn from 1F9M.pdb.]

stroma to compensate for the influx of protons. Carbamate formation is favored at alkaline pH. CO_2 adds to a deprotonated form of lysine 201 of rubisco, and Mg^{2+} ion binds to the carbamate to generate the active form of the enzyme. Thus, light leads to the generation of regulatory signals as well as ATP and NADPH.

Thioredoxin plays a key role in regulating the Calvin cycle

Light-driven reactions lead to electron transfer from water to ferredoxin and, eventually, to NADPH. The presence of reduced ferredoxin and NADPH are good signals that conditions are right for biosynthesis. One way in which this information is conveyed to biosynthetic enzymes is by *thioredoxin,* a 12-kDa protein containing neighboring cysteine residues that cycle between a reduced sulfhydryl and an oxidized disulfide form (Figure 20.14). The reduced form of thioredoxin activates many biosynthetic enzymes, including the chloroplast ATP synthase (p. 579), by reducing disulfide bridges that control their activity. Reduced thioredoxin also inhibits several degradative enzymes by the same means (Table 20.1). In chloroplasts, oxidized thioredoxin is reduced by ferredoxin in a reaction catalyzed by *ferredoxin–thioredoxin reductase* (p. 580). Thus, *the activities of the light and dark reactions of photosynthesis are coordinated through electron transfer from reduced ferredoxin to thioredoxin and then to component enzymes containing regulatory disulfide bonds* (Figure 20.15). We shall return to thioredoxin when we consider the reduction of ribonucleotides (Section 25.3).

NADPH is a signal molecule that activates two biosynthetic enzymes, phosphoribulokinase and glyceraldehyde 3-phosphate dehydrogenase. In the dark, these enzymes are inhibited by association with an 8.5-kDa protein called *CP12,* an intrinsically disordered protein (Section 2.6). NADPH disrupts this association by promoting the formation of two disulfide bonds in CP12, leading to the release of the active enzymes.

The C_4 pathway of tropical plants accelerates photosynthesis by concentrating carbon dioxide

The oxygenase activity of rubisco presents a biochemical challenge to tropical plants because the oxygenase activity increases more rapidly with temperature than does the carboxylase activity. How, then, do plants that grow in hot climates, such as sugar cane, prevent very high rates of wasteful photorespiration? Their solution to this problem is to achieve a high local concentration of CO_2 at the site of the Calvin cycle in their photosynthetic cells. The essence of this process, which was elucidated by Marshall Davidson Hatch and C. Roger Slack, is that *four-carbon (C_4) compounds such as oxaloacetate and malate carry CO_2 from mesophyll cells, which are in*

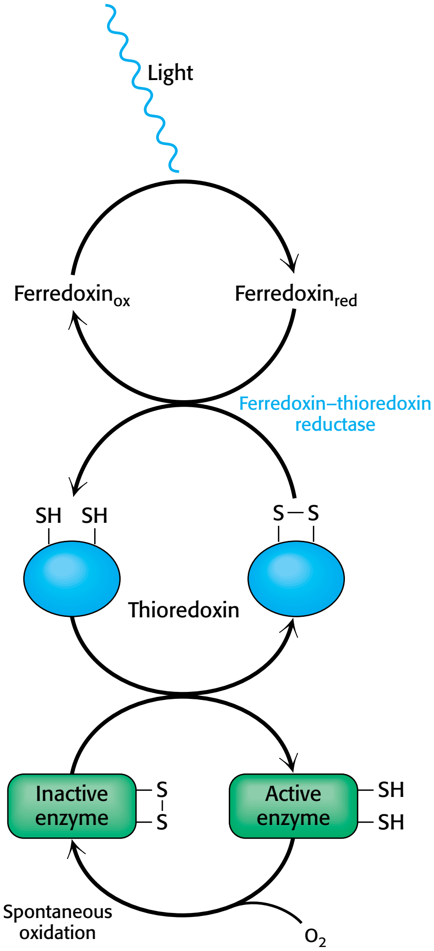


FIGURE 20.15 Enzyme activation by thioredoxin. Reduced thioredoxin activates certain Calvin cycle enzymes by cleaving regulatory disulfide bonds.

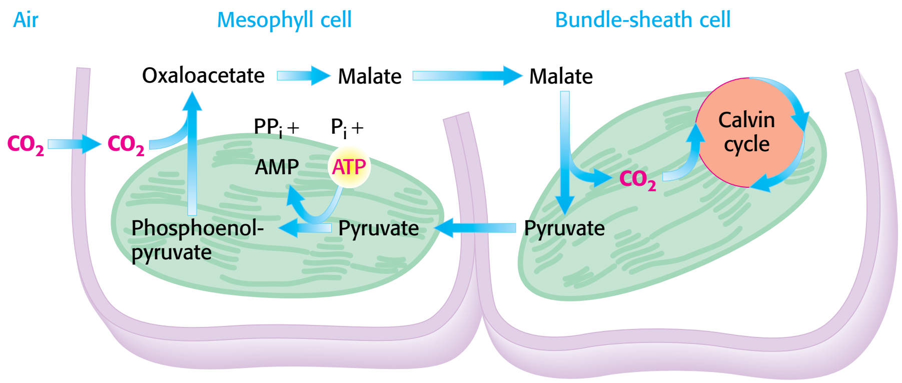


FIGURE 20.16 C_4 pathway. Carbon dioxide is concentrated in bundle-sheath cells by the expenditure of ATP in mesophyll cells.

contact with air, to bundle-sheath cells, which are the major sites of photosynthesis (Figure 20.16). The decarboxylation of the four-carbon compound in a bundle-sheath cell maintains a high concentration of CO_2 at the site of the Calvin cycle. The three-carbon product returns to the mesophyll cell for another round of carboxylation.

The *C_4 pathway* for the transport of CO_2 starts in a mesophyll cell with the condensation of CO_2 and phosphoenolpyruvate to form *oxaloacetate* in a reaction catalyzed by *phosphoenolpyruvate carboxylase*. In some species, oxaloacetate is converted into malate by an $NADP^+$-linked malate dehydrogenase. Malate enters the bundle-sheath cell and is oxidatively decarboxylated within the chloroplasts by an $NADP^+$—linked malate dehydrogenase. The released CO_2 enters the Calvin cycle in the usual way by condensing with ribulose 1,5-bisphosphate. Pyruvate formed in this decarboxylation reaction returns to the mesophyll cell. Finally, phosphoenolpyruvate is formed from pyruvate by *pyruvate-P_i dikinase*.

The net reaction of this C_4 pathway is

$$\begin{aligned} &CO_2 \text{ (in mesophyll cell)} + ATP + 2\,H_2O \longrightarrow \\ &\qquad CO_2 \text{ (in bundle-sheath cell)} + AMP + 2\,P_i + 2\,H^+ \end{aligned}$$

Thus, *the energetic equivalent of two ATP molecules is consumed in transporting CO_2 to the chloroplasts of the bundle-sheath cells*. In essence, this process is active transport: the pumping of CO_2 into the bundle-sheath cell is driven by the hydrolysis of one molecule of ATP to one molecule of AMP and two molecules of orthophosphate. The CO_2 concentration can be 20-fold as great in the bundle-sheath cells as in the mesophyll cells.

When the C_4 pathway and the Calvin cycle operate together, the net reaction is

$$\begin{aligned} &6\,CO_2 + 30\,ATP + 12\,NADPH + 24\,H_2O \longrightarrow \\ &\qquad C_6H_{12}O_6 + 30\,ADP + 30\,P_i + 12\,NADP^+ + 18\,H^+ \end{aligned}$$

Note that 30 molecules of ATP are consumed per hexose molecule formed when the C_4 pathway delivers CO_2 to the Calvin cycle, in contrast with 18 molecules of ATP per hexose molecule in the absence of the C_4 pathway. The high concentration of CO_2 in the bundle-sheath cells of C_4 plants, which is the result of the expenditure of the additional 12 molecules of ATP, is critical for their rapid photosynthetic rate, because CO_2 is limiting when light is abundant. A high CO_2 concentration also minimizes the energy loss caused by photorespiration.

Tropical plants with a C_4 pathway do little photorespiration because the high concentration of CO_2 in their bundle-sheath cells accelerates the carboxylase reaction relative to the oxygenase reaction. This effect is especially important at higher temperatures. The geographic distribution of plants having this pathway (C_4 plants) and those lacking it (C_3 plants) can now be understood in molecular terms. *C_4 plants* have the advantage in a hot environment and under high illumination, which accounts for their prevalence in the tropics. *C_3 plants,* which consume only 18 molecules of ATP per hexose molecule formed in the absence of photorespiration (compared with 30 molecules of ATP for C_4 plants), are more efficient at temperatures lower than about 28°C, and so they predominate in temperate environments.

Rubisco is present in bacteria, eukaryotes, and even archaea, although other photosynthetic components have not been found in archaea. Thus, rubisco emerged early in evolution, when the atmosphere was rich in CO_2 and almost devoid of O_2. The enzyme was not originally selected to operate in an environment like the present one, which is almost devoid of CO_2 and rich in O_2. Photorespiration became significant about 600 million years ago, when the CO_2 concentration fell to present levels. The C_4 pathway is thought to have evolved in response no more than 30 million years ago and possibly as recently as 7 million years ago. It is interesting that none of the enzymes are unique to C_4 plants, suggesting that this pathway made use of already existing enzymes.

Crassulacean acid metabolism permits growth in arid ecosystems

Many plants, including some that grow in hot, dry climates, keep the stomata of their leaves closed in the heat of the day to prevent water loss (Figure 20.17). As a consequence, CO_2 cannot be absorbed during the daylight hours, when it is needed for glucose synthesis. Rather, CO_2 enters the leaf when the stomata open at the cooler temperatures of night. To store the CO_2 until it can be used during the day, such plants make use of an adaptation called *crassulacean acid metabolism* (CAM), named after the genus *Crassulacea* (the succulents). Carbon dioxide is fixed by the C_4 pathway into malate, which is stored in vacuoles. During the day, malate is decarboxylated and the CO_2 becomes available to the Calvin cycle. In contrast with C_4 plants, CAM plants separate CO_2 accumulation from CO_2 utilization temporally rather than spatially.

Although CAM plants do prevent water loss, the use of malate as the sole source of CO_2 comes at a metabolic cost. Because the malate storage is limited, CAM plants cannot generate CO_2 as rapidly as it can be imported by C_3 and C_4 plants. Consequently, the growth rate of CAM plants is slower than that of C_3 and C_4 plants. The saguaro cactus, which can live up to 200 years and reach a height of 60 feet, can take 15 years to grow to only 1 foot in height.

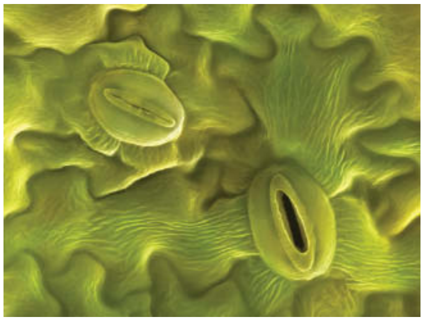


FIGURE 20.17 Scanning electron micrograph of an open stoma and a closed stoma. [Power and Syred/Science Photo Library.]

20.3 The Pentose Phosphate Pathway Generates NADPH and Synthesizes Five-Carbon Sugars

Photosynthetic organisms can use the light reactions for generation of some NADPH. Another pathway, present in all organisms, meets the NADPH needs of nonphotosynthetic organisms and of the nonphotosynthetic tissues in plants. The *pentose phosphate pathway,* which occurs in the cytoplasm, is a crucial source of NADPH for use in reductive biosynthesis (Table 20.2) as well as for protection against oxidative stress. This pathway consists of two phases: (1) the oxidative generation of NADPH and (2) the nonoxidative interconversion of sugars (Figure 20.18). In the oxidative phase, NADPH is generated when glucose 6-phosphate is oxidized to ribulose 5-phosphate.

$$\text{Glucose 6-phosphate} + 2\,\text{NADP}^+ + H_2O \longrightarrow \text{ribulose 5-phosphate} + 2\,\text{NADPH} + 2\,H^+ + CO_2$$

TABLE 20.2 Pathways requiring NADPH

Synthesis
Fatty acid biosynthesis
Cholesterol biosynthesis
Neurotransmitter biosynthesis
Nucleotide biosynthesis
Detoxification
Reduction of oxidized glutathione
Cytochrome P450 monooxygenases

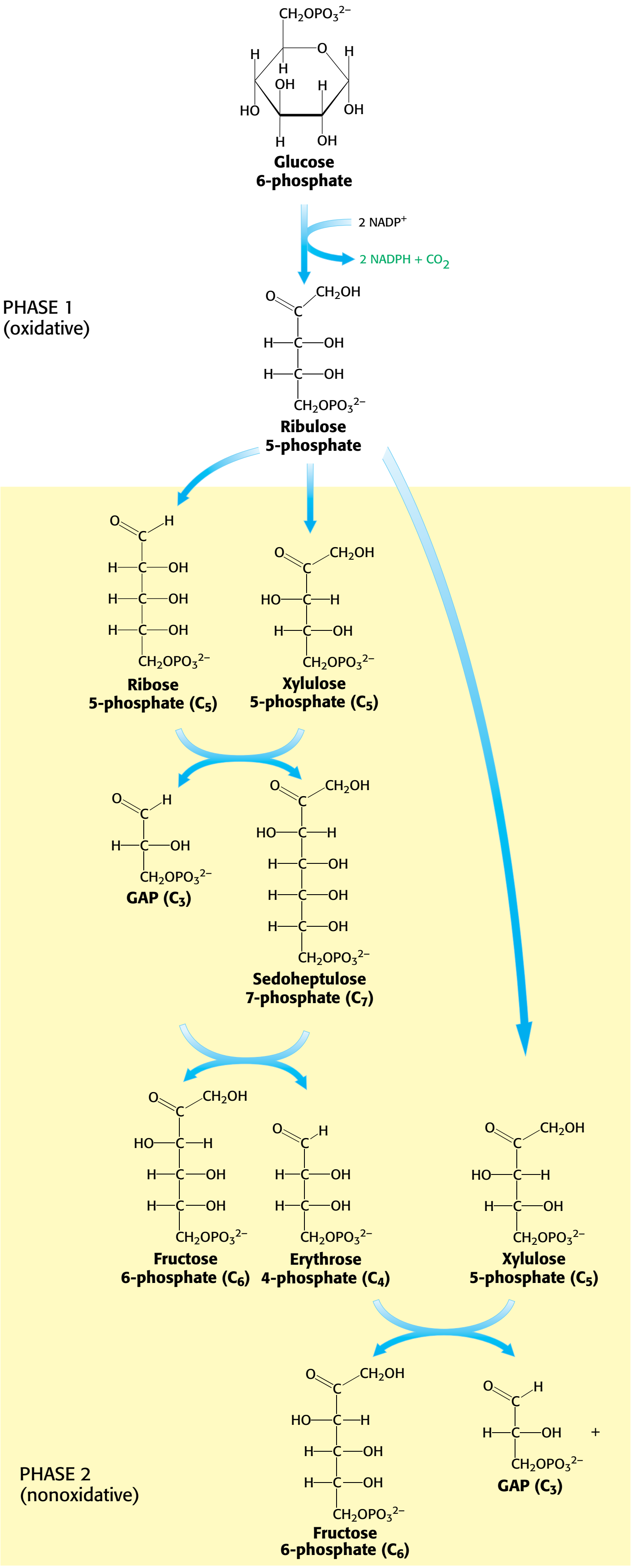

Glucose 6-phosphate
2 NADP+
2 NADPH + CO2
PHASE 1 (oxidative)
Ribulose 5-phosphate
Ribose 5-phosphate (C5)
Xylulose 5-phosphate (C5)
GAP (C3)
Sedoheptulose 7-phosphate (C7)
Fructose 6-phosphate (C6)
Erythrose 4-phosphate (C4)
Xylulose 5-phosphate (C5)
Fructose 6-phosphate (C6)
GAP (C3)
PHASE 2 (nonoxidative)

FIGURE 20.18 Pentose phosphate pathway. The pathway consists of (1) an oxidative phase that generates NADPH and (2) a nonoxidative phase that interconverts phosphorylated sugars.

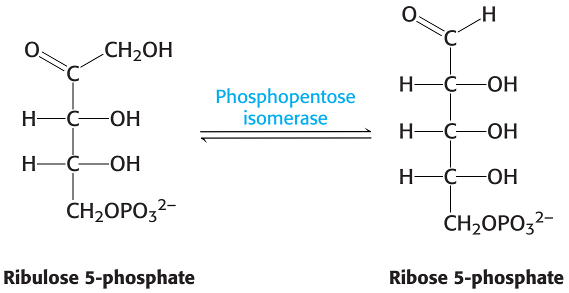
In the nonoxidative phase, the pathway catalyzes the interconversion of three-, four-, five-, six-, and seven-carbon sugars in a series of nonoxidative reactions. Excess five-carbon sugars may be converted into intermediates of the glycolytic pathway. All these reactions take place in the cytoplasm. These interconversions rely on the same reactions that lead to the regeneration of ribulose 1,5-bisphosphate in the Calvin cycle.

Two molecules of NADPH are generated in the conversion of glucose 6-phosphate into ribulose 5-phosphate

The oxidative phase of the pentose phosphate pathway starts with the dehydrogenation of glucose 6-phosphate at carbon 1, a reaction catalyzed by *glucose 6-phosphate dehydrogenase* (Figure 20.19). This enzyme is highly specific for $NADP^+$; the K_M for NAD^+ is about a thousand times as great as that for $NADP^+$. The product is *6-phosphoglucono-δ-lactone,* which is an intramolecular ester between the C-1 carboxyl group and the C-5 hydroxyl group. The next step is the hydrolysis of 6-phosphoglucono-δ-lactone by a specific *lactonase* to give *6-phosphogluconate*. This six-carbon sugar acid is then oxidatively decarboxylated by *6-phosphogluconate dehydrogenase* to yield *ribulose 5-phosphate*. $NADP^+$is again the electron acceptor.

The pentose phosphate pathway and glycolysis are linked by transketolase and transaldolase

The preceding reactions yield two molecules of NADPH and one molecule of ribulose 5-phosphate for each molecule of glucose 6-phosphate oxidized. The ribulose 5-phosphate is subsequently isomerized to ribose 5-phosphate by phosphopentose isomerase.


Phosphopentose isomerase
Ribulose 5-phosphate
Ribose 5-phosphate

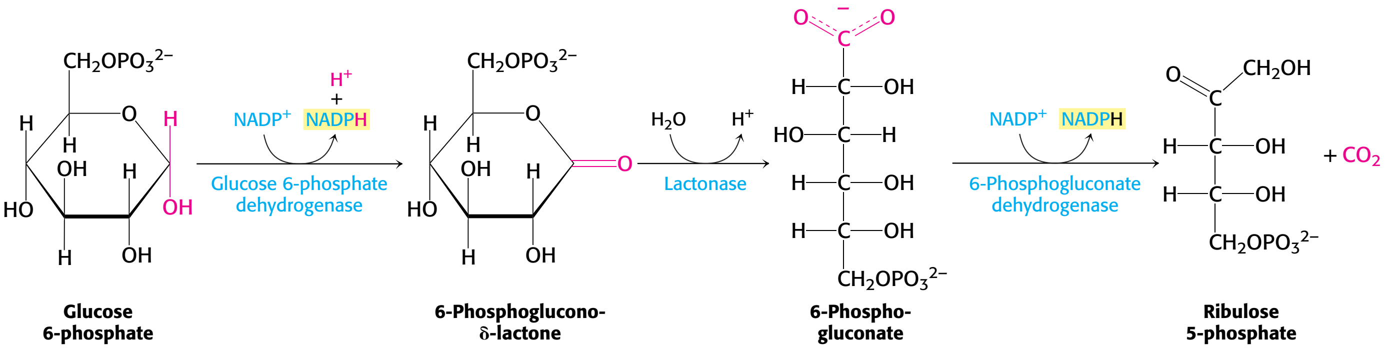

NADP⁺
H⁺ + NADPH
Glucose 6-phosphate dehydrogenase
H₂O
H⁺
Lactonase
NADP⁺
NADPH
6-Phosphogluconate dehydrogenase
+ CO₂
Glucose 6-phosphate
6-Phosphoglucono-δ-lactone
6-Phosphogluconate
Ribulose 5-phosphate

FIGURE 20.19 Oxidative phase of the pentose phosphate pathway. Glucose 6-phosphate is oxidized to 6-phosphoglucono-δ-lactone to generate one molecule of NADPH. The lactone product is hydrolyzed to 6-phosphogluconate, which is oxidatively decarboxylated to ribulose 5-phosphate with the generation of a second molecule of NADPH.

Ribose 5-phosphate and its derivatives are components of RNA and DNA, as well as of ATP, NADH, FAD, and coenzyme A. Although ribose 5-phosphate is a precursor to many biomolecules, many cells need NADPH for reductive biosyntheses much more than they need ribose 5-phosphate for incorporation into nucleotides. For instance, adipose tissue, the liver, and mammary glands require large amounts of NADPH for fatty acid synthesis (Chapter 22). In these cases, ribose 5-phosphate is converted into the glycolytic intermediates glyceraldehyde 3-phosphate and fructose 6-phosphate by *transketolase* and *transaldolase. These enzymes create a reversible link between the pentose phosphate pathway and glycolysis by catalyzing these three successive reactions.*

$$C_5 + C_5 \xrightleftharpoons{\text{Transketolase}} C_3 + C_7$$

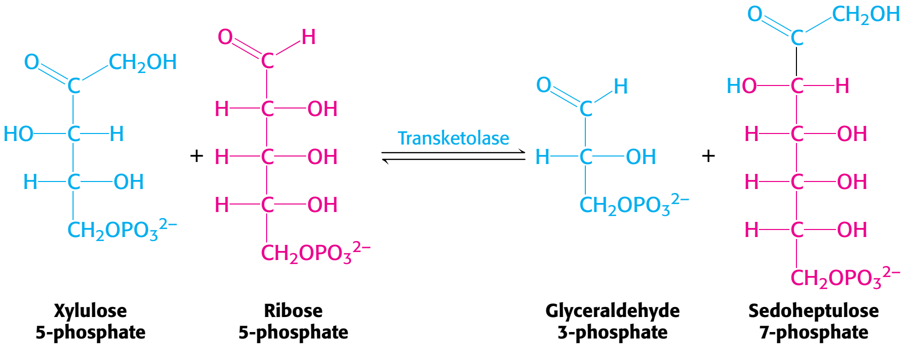
$$C_3 + C_7 \xrightleftharpoons{\text{Transaldolase}} C_6 + C_4$$

$$C_4 + C_5 \xrightleftharpoons{\text{Transketolase}} C_6 + C_3$$

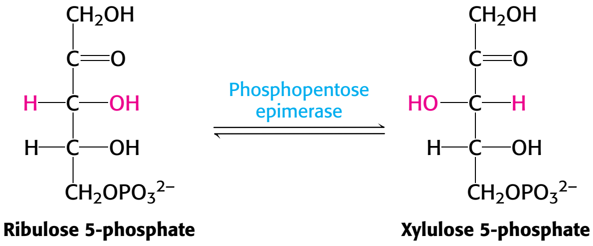
The net result of these reactions is the *formation of two hexoses and one triose from three pentoses:*

$$3\,C_5 \rightleftharpoons 2\,C_6 + C_3$$

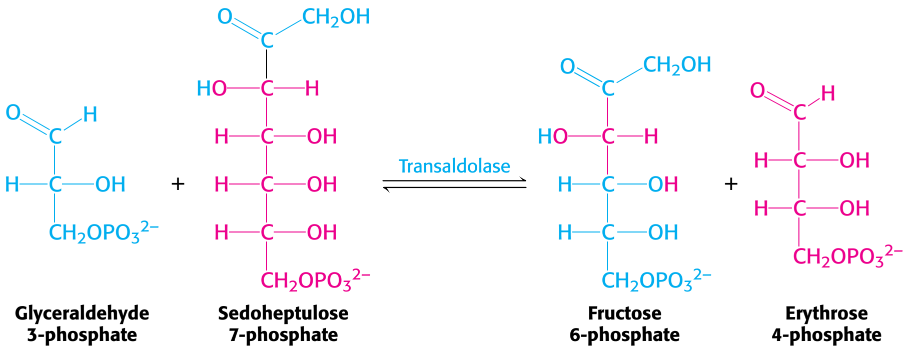
The first of the three reactions linking the pentose phosphate pathway and glycolysis is the formation of *glyceraldehyde 3-phosphate* and *sedoheptulose 7-phosphate* from two pentoses.


Transketolase
Xylulose 5-phosphate
Ribose 5-phosphate
Glyceraldehyde 3-phosphate
Sedoheptulose 7-phosphate

The donor of the two-carbon unit in this reaction is xylulose 5-phosphate, an epimer of ribulose 5-phosphate. A ketose is a substrate of transketolase only if its hydroxyl group at C-3 has the configuration of xylulose rather than ribulose. Ribulose 5-phosphate is converted into the appropriate epimer for the transketolase reaction by *phosphopentose epimerase* in the reverse reaction of that which takes place in the Calvin cycle.

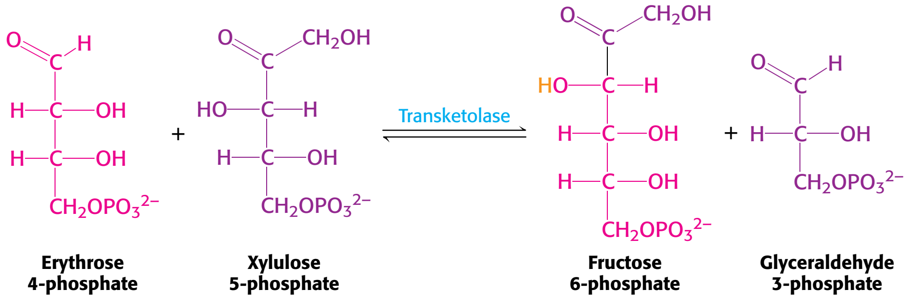

Phosphopentose epimerase
Ribulose 5-phosphate
Xylulose 5-phosphate

Glyceraldehyde 3-phosphate and sedoheptulose 7-phosphate generated by the transketolase then react to form *fructose 6-phosphate* and *erythrose 4-phosphate*.


Transaldolase
Glyceraldehyde 3-phosphate
Sedoheptulose 7-phosphate
Fructose 6-phosphate
Erythrose 4-phosphate

This synthesis of a four-carbon sugar and a six-carbon sugar is catalyzed by *transaldolase*.

In the third reaction, transketolase catalyzes the synthesis of *fructose 6-phosphate* and *glyceraldehyde 3-phosphate* from erythrose 4-phosphate and xylulose 5-phosphate.


Transketolase
Erythrose 4-phosphate
Xylulose 5-phosphate
Fructose 6-phosphate
Glyceraldehyde 3-phosphate

The sum of these reactions is

$$\text{2 Xylulose 5-phosphate} + \text{ribose 5-phosphate} \rightleftharpoons \text{2 fructose 6-phosphate} + \text{glyceraldehyde 3-phosphate}$$

Xylulose 5-phosphate can be formed from ribose 5-phosphate by the sequential action of phosphopentose isomerase and phosphopentose epimerase, and so the net reaction starting from ribose 5-phosphate is

$$\text{3 Ribose 5-phosphate} \rightleftharpoons \text{2 fructose 6-phosphate} + \text{glyceraldehyde 3-phosphate}$$

Thus, *excess ribose 5-phosphate formed by the pentose phosphate pathway can be completely converted into glycolytic intermediates*. Moreover, any ribose

TABLE 20.3 Pentose phosphate pathway

Reaction	Enzyme
Oxidative phase	
Glucose 6-phosphate + NADP$^+$ $\longrightarrow$ 6-phosphoglucono-δ-lactone + NADPH + H$^+$	Glucose 6-phosphate dehydrogenase
6-Phosphoglucono-δ-lactone + H_2O $\longrightarrow$ 6-phosphogluconate + H$^+$	Lactonase
6-Phosphogluconate + NADP$^+$ $\longrightarrow$ ribulose 5-phosphate + CO_2 + NADPH + H$^+$	6-Phosphogluconate dehydrogenase
Nonoxidative Phase	
Ribulose 5-phosphate $\rightleftharpoons$ ribose 5-phosphate	Phosphopentose isomerase
Ribulose 5-phosphate $\rightleftharpoons$ xylulose 5-phosphate	Phosphopentose epimerase
Xylulose 5-phosphate + ribose 5-phosphate $\rightleftharpoons$ sedoheptulose 7-phosphate + glyceraldehyde 3-phosphate	Transketolase
Sedoheptulose 7-phosphate + glyceraldehyde 3-phosphate $\rightleftharpoons$ fructose 6-phosphate + erythrose 4-phosphate	Transaldolase
Xylulose 5-phosphate + erythrose 4-phosphate $\rightleftharpoons$ fructose 6-phosphate + glyceraldehyde 3-phosphate	Transketolase

ingested in the diet can be processed into glycolytic intermediates by this pathway. It is evident that the carbon skeletons of sugars can be extensively rearranged to meet physiological needs (Table 20.3).

Mechanism: Transketolase and transaldolase stabilize carbanionic intermediates by different mechanisms

The reactions catalyzed by transketolase and transaldolase are distinct, yet similar in many ways. One difference is that transketolase transfers a two-carbon unit, whereas transaldolase transfers a three-carbon unit. Each of these units is transiently attached to the enzyme in the course of the reaction and thus these enzymes are examples of double displacement reactions (Section 8.4).

Transketolase reaction. Transketolase contains a tightly bound thiamine pyrophosphate as its prosthetic group. The enzyme transfers a two-carbon glycoaldehyde from a ketose donor to an aldose acceptor. The site of the addition of the two-carbon unit is the thiazole ring of TPP. Transketolase is homologous to the E_1 subunit of the pyruvate dehydrogenase complex (Section 17.1) and the reaction mechanism is similar (Figure 20.20).

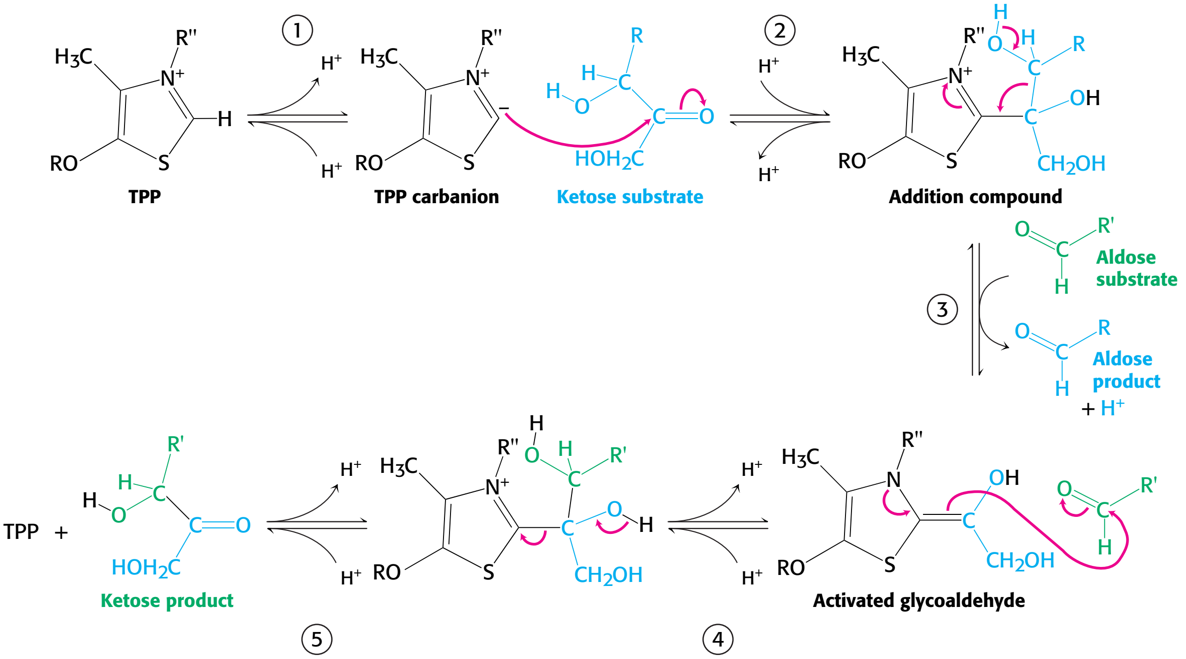


FIGURE 20.20 Transketolase mechanism. (1) Thiamine pyrophosphate (TPP) ionizes to form a carbanion. (2) The carbanion of TPP attacks the ketose substrate. (3) Cleavage of a carbon–carbon bond frees the aldose product and leaves a two-carbon fragment joined to TPP. (4) This activated glycoaldehyde intermediate attacks the aldose substrate to form a new carbon–carbon bond. (5) The ketose product is released, freeing the TPP for the next reaction cycle.

The reaction proceeds as follows:

1. The C-2 carbon atom of bound TPP readily ionizes to give a *carbanion*.
2. The negatively charged carbon atom of this reactive intermediate attacks the carbonyl group of the ketose substrate.
3. The resulting addition compound releases the aldose product to yield an *activated glycoaldehyde unit*.
4. The positively charged nitrogen atom in the thiazole ring acts as an *electron sink* in the development of this activated intermediate. The carbonyl group of a suitable aldose acceptor then condenses with the activated glycoaldehyde unit to form a new ketose.
5. The ketose is then released from the enzyme.

Transaldolase reaction. Transaldolase transfers a three-carbon *dihydroxyacetone* unit from a ketose donor to an aldose acceptor. Transaldolase, in contrast to transketolase, does not contain a prosthetic group. Rather, *a Schiff base is formed between the carbonyl group of the ketose substrate and the ε-amino group of a lysine residue at the active site of the enzyme* (Figure 20.21). This kind of covalent enzyme–substrate intermediate is like that formed in fructose 1,6-bisphosphate aldolase in the glycolytic pathway (Section 16.1) and, indeed, the enzymes are homologous. (1) The reaction begins with the formation of the Schiff base. (2) The Schiff base becomes protonated and the bond between C-3 and C-4 is split. (3) Upon reprotonation, the aldose product is released leaving a three-carbon fragment bound to the enzyme. The negative charge on the Schiff-base carbanion moiety is stabilized by resonance (Figure 20.22). The positively charged nitrogen atom of the protonated Schiff base acts as an electron sink. (4) The Schiff-base adduct is stable until a suitable aldose becomes bound. (5) The dihydroxyacetone moiety then reacts with the carbonyl group of the aldose.

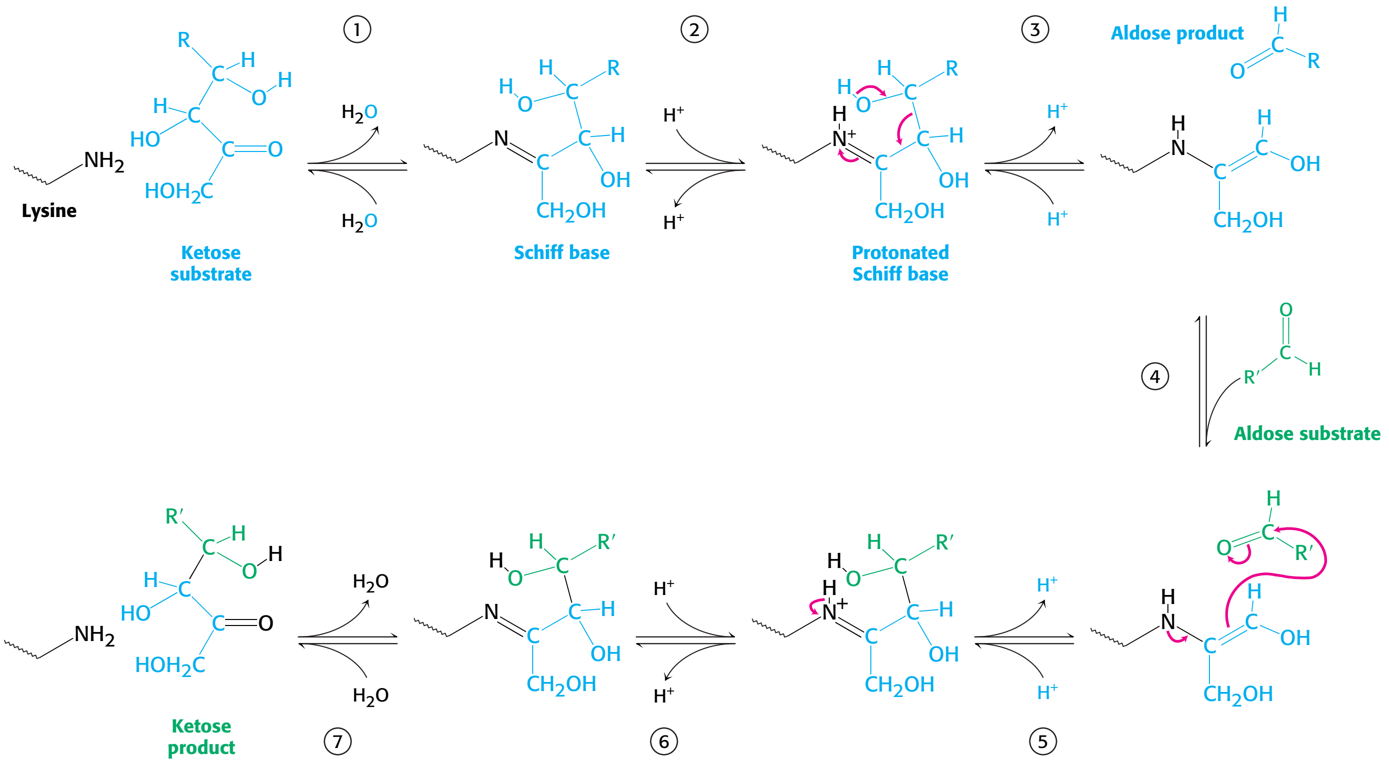


FIGURE 20.21 Transaldolase mechanism. (1) The reaction begins with the formation of a Schiff base between a lysine residue in transaldolase and the ketose substrate. (2) Protonation of the Schiff base occurs. (3) Deprotonation leads to the release of the aldose product, leaving a three-carbon fragment attached to the lysine residue. (4) This intermediate adds to the aldose substrate. (5) Protonation occurs, forming a new carbon–carbon bond. (6) Subsequent deprotonation and (7) hydrolysis of the Schiff base releases the ketose product from the lysine side chain, completing the reaction cycle.

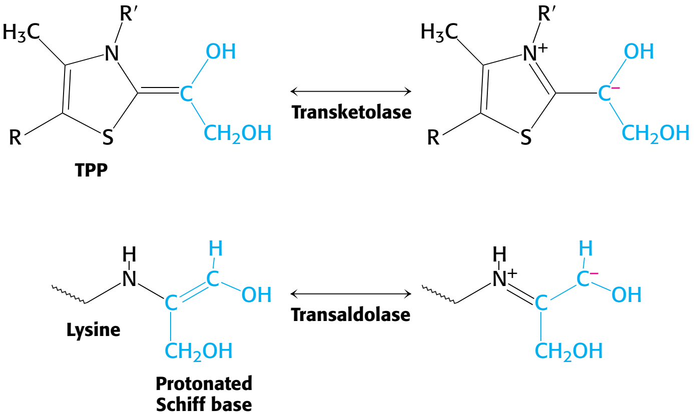


FIGURE 20.22 Carbanion intermediates. For transketolase and transaldolase, a carbanion intermediate is stabilized by resonance. In transketolase, TPP stabilizes this intermediate; in transaldolase, a protonated Schiff base plays this role.

Protonation allows the formation of a new carbon–carbon bond. (6) Subsequent deprotonation followed by (7) hydrolysis of the Schiff base releases the ketose product. The nitrogen atom of the protonated Schiff base plays the same role in transaldolase as the thiazole-ring nitrogen atom does in transketolase. In each enzyme, a group within an intermediate reacts like a carbanion in attacking a carbonyl group to form a new carbon–carbon bond. In each case, the charge on the carbanion is stabilized by resonance (Figure 20.22).

20.4 The Metabolism of Glucose 6-Phosphate by the Pentose Phosphate Pathway Is Coordinated with Glycolysis

Glucose 6-phosphate is metabolized by both the glycolytic pathway (Chapter 16) and the pentose phosphate pathway. How is the processing of this important metabolite partitioned between these two metabolic routes? The cytoplasmic concentration of $NADP^+$ plays a key role in determining the fate of glucose 6-phosphate.

The rate of the pentose phosphate pathway is controlled by the level of $NADP^+$

The first reaction in the oxidative branch of the pentose phosphate pathway, the dehydrogenation of glucose 6-phosphate, is essentially irreversible. In fact, this reaction is rate limiting under physiological conditions and serves as the control site for the oxidative branch of the pathway. The most important regulatory factor is the level of $NADP^+$. Low levels of $NADP^+$ reduce the dehydrogenation of glucose 6-phosphate because it is needed as the electron acceptor. The effect of low levels of $NADP^+$ is intensified by the fact that NADPH competes with $NADP^+$ in binding to the enzyme. The ratio of $NADP^+$ to NADPH in the cytoplasm of a liver cell from a well-fed rat is about 0.014. The marked effect of the $NADP^+$ level on the rate of the oxidative phase ensures that NADPH is not generated unless the supply needed for reductive biosyntheses or protection against oxidative stress is low. The nonoxidative phase of the pentose phosphate pathway is controlled primarily by the availability of substrates.

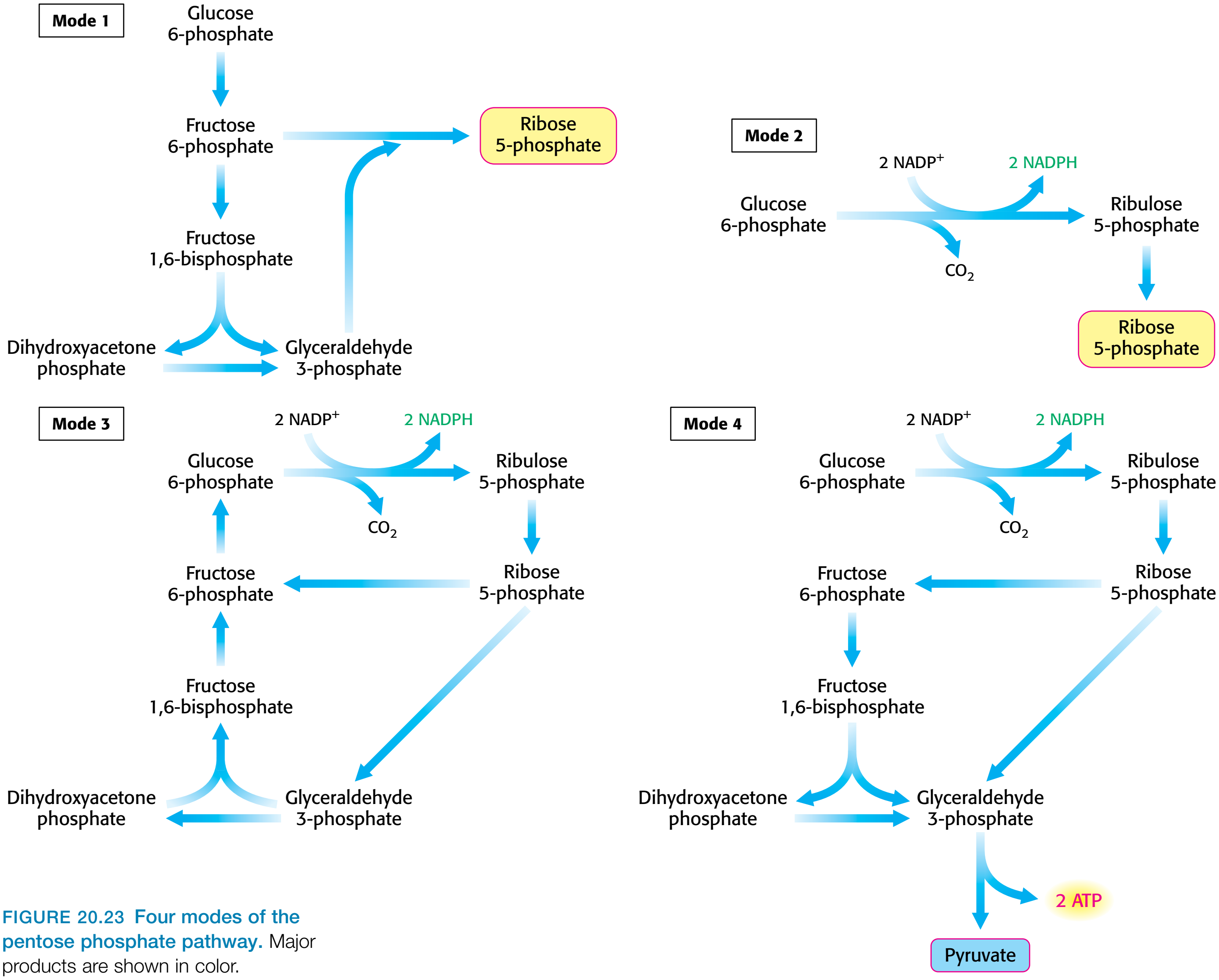

Mode 1
Glucose 6-phosphate
Fructose 6-phosphate
Ribose 5-phosphate
Fructose 1,6-bisphosphate
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
Mode 2
2 NADP+
2 NADPH
Glucose 6-phosphate
Ribulose 5-phosphate
CO2
Ribose 5-phosphate
Mode 3
2 NADP+
2 NADPH
Glucose 6-phosphate
Ribulose 5-phosphate
CO2
Fructose 6-phosphate
Ribose 5-phosphate
Fructose 1,6-bisphosphate
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
Mode 4
2 NADP+
2 NADPH
Glucose 6-phosphate
Ribulose 5-phosphate
CO2
Fructose 6-phosphate
Ribose 5-phosphate
Fructose 1,6-bisphosphate
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
2 ATP
Pyruvate

FIGURE 20.23 Four modes of the pentose phosphate pathway. Major products are shown in color.

The flow of glucose 6-phosphate depends on the need for NADPH, ribose 5-phosphate, and ATP

We can grasp the intricate interplay between glycolysis and the pentose phosphate pathway by examining the metabolism of glucose 6-phosphate in four different metabolic situations (Figure 20.23).

Mode 1. *Much more ribose 5-phosphate than NADPH is required*. For example, rapidly dividing cells need ribose 5-phosphate for the synthesis of nucleotide precursors of DNA. Most of the glucose 6-phosphate is converted into fructose 6-phosphate and glyceraldehyde 3-phosphate by the glycolytic pathway. Transaldolase and transketolase then convert two molecules of fructose 6-phosphate and one molecule of glyceraldehyde 3-phosphate into three molecules of ribose 5-phosphate by a reversal of the reactions described earlier. The stoichiometry of mode 1 is

$$5 \text{ Glucose 6-phosphate} + \text{ATP} \rightarrow 6 \text{ ribose 5-phosphate} + \text{ADP} + 2\,\text{H}^+$$

Mode 2. *The needs for NADPH and for ribose 5-phosphate are balanced.* Under these conditions, glucose 6-phosphate is processed to one molecule of ribulose 5-phosphate while generating two molecules of NADPH. Ribulose

TABLE 20.4 Tissues with active pentose phosphate pathways

Tissue	Function
Adrenal gland	Steroid synthesis
Liver	Fatty acid and cholesterol synthesis
Testes	Steroid synthesis
Adipose tissue	Fatty acid synthesis
Ovary	Steroid synthesis
Mammary gland	Fatty acid synthesis
Red blood cells	Maintenance of reduced glutathione

5-phosphate is then converted into ribose 5-phosphate. The stoichiometry of mode 2 is

$$\text{Glucose 6-phosphate} + 2\ \text{NADP}^+ + \text{H}_2\text{O} \longrightarrow \text{ribose 5-phosphate} + 2\ \text{NADPH} + 2\ \text{H}^+ + \text{CO}_2$$

Mode 3. *Much more NADPH than ribose 5-phosphate is required.* For example, adipose tissue requires a high level of NADPH for the synthesis of fatty acids (Table 20.4). In this case, glucose 6-phosphate is completely oxidized to CO_2. Three groups of reactions are active in this situation. First, the oxidative phase of the pentose phosphate pathway forms two molecules of NADPH and one molecule of ribulose 5-phosphate. Then, ribulose 5-phosphate is converted into fructose 6-phosphate and glyceraldehyde 3-phosphate by transketolase and transaldolase. Finally, glucose 6-phosphate is resynthesized from fructose 6-phosphate and glyceraldehyde 3-phosphate by the gluconeogenic pathway. The stoichiometries of these three sets of reactions are

$$6\ \text{Glucose 6-phosphate} + 12\ \text{NADP}^+ + 6\ \text{H}_2\text{O} \longrightarrow 6\ \text{Ribose 5-phosphate} + 12\ \text{NADPH} + 12\ \text{H}^+ + 6\text{CO}_2$$

$$6\ \text{Ribose 5-phosphate} \longrightarrow 4\ \text{Fructose 6-phosphate} + 2\ \text{glyceraldehyde 3-phosphate}$$

$$4\ \text{Fructose 6-phosphate} + 2\ \text{glyceraldehyde 3-phosphate} + \text{H}_2\text{O} \longrightarrow 5\text{-Glucose 6-phosphate} + \text{P}_\text{i}$$

The sum of the mode 3 reactions is

$$\text{Glucose 6-phosphate} + 12\ \text{NADP}^+ + 7\ \text{H}_2\text{O} \longrightarrow 6\ \text{CO}_2 + 12\ \text{NADPH} + 12\ \text{H}^+ + \text{P}_\text{i}$$

Thus, *the equivalent of glucose 6-phosphate can be completely oxidized to CO_2 with the concomitant generation of NADPH.* In essence, ribose 5-phosphate produced by the pentose phosphate pathway is recycled into glucose 6-phosphate by transketolase, transaldolase, and some of the enzymes of the gluconeogenic pathway.

Mode 4. *Both NADPH and ATP are required.* Alternatively, ribulose 5-phosphate formed from glucose 6-phosphate can be converted into pyruvate. Fructose 6-phosphate and glyceraldehyde 3-phosphate derived from ribose 5-phosphate enter the glycolytic pathway rather than reverting to glucose 6-phosphate. In this mode, *ATP and NADPH are concomitantly generated, and five of the six carbons of glucose 6-phosphate emerge in pyruvate.*

$$3\ \text{Glucose 6-phosphate} + 6\ \text{NADP}^+ + 5\ \text{NAD}^+ + 5\ \text{P}_\text{i} + 8\ \text{ADP} \longrightarrow 5\ \text{pyruvate} + 3\ \text{CO}_2 + 6\ \text{NADPH} + 5\ \text{NADH} + 8\ \text{ATP} + 2\ \text{H}_2\text{O} + 8\ \text{H}^+$$

Pyruvate formed by these reactions can be oxidized to generate more ATP or it can be used as a building block in a variety of biosyntheses.

The pentose phosphate pathway is required for rapid cell growth

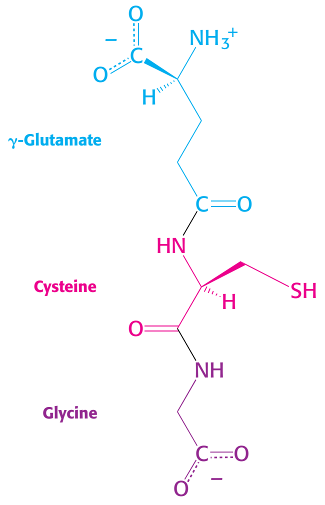
Rapidly dividing cells, such as cancer cells, require ribose 5-phosphate for nucleic acid synthesis and NADPH for fatty acid synthesis, which in turn is required to form membrane lipids (Section 26.1). Recall that rapidly dividing cells switch to aerobic glycolysis to meet their ATP needs (Section 16.2). Glucose 6-phosphate and glycolytic intermediates are then used to generate NADPH and ribose 5-phosphate using the pentose phosphate pathway as described in Modes 1 and 3. The diversion of glycolytic intermediates into the nonoxidative phase is facilitated by the expression of the gene for a pyruvate kinase isozyme, PKM. PKM has a low catalytic activity, creating a bottleneck in the glycolytic pathway. Glycolytic intermediates accumulate and are then used by the pentose phosphate pathway to synthesize NADPH and ribose 5-phosphate. The shunting of phosphorylated intermediates into the nonoxidative phase of the pentose phosphate pathway is further enabled by the inhibition of triose phosphate isomerase by phosphoenolpyruvate, the substrate of PKM.

Through the looking-glass: The Calvin cycle and the pentose phosphate pathway are mirror images

The complexities of the Calvin cycle and the pentose phosphate pathway are easier to comprehend if we consider them as functional mirror images of each other. The Calvin cycle begins with the fixation of CO_2 and proceeds to use NADPH in the synthesis of glucose. The pentose phosphate pathway begins with the oxidation of a glucose-derived carbon atom to CO_2 and concomitantly generates NADPH. The regeneration phase of the Calvin cycle converts C_6 and C_3 molecules back into the starting material—the C_5 molecule ribulose 1,5-bisphosphate. The pentose phosphate pathway converts a C_5 molecule, ribulose 5-phosphate, into C_6 and C_3 intermediates of the glycolytic pathway. Not surprisingly, in photosynthetic organisms, many enzymes are common to the two pathways. We see the economy of evolution: the use of identical enzymes for similar reactions with different ends.

20.5 Glucose 6-Phosphate Dehydrogenase Plays a Key Role in Protection Against Reactive Oxygen Species

The NADPH generated by the pentose phosphate pathway plays a vital role in protecting the cells from reactive oxygen species (ROS). Reactive oxygen species generated in oxidative metabolism inflict damage on all classes of macromolecules and can ultimately lead to cell death. Indeed, ROS are implicated in a number of human diseases (Table 18.3). Reduced *glutathione* (GSH), a tripeptide with a free sulfhydryl group, combats oxidative stress by reducing ROS to harmless forms. Its task accomplished, the glutathione is now in the oxidized form (GSSG) and must be reduced to regenerate GSH. The reducing power is supplied by the NADPH generated by glucose 6-phosphate dehydrogenase in the pentose phosphate pathway. Indeed, cells with reduced levels of glucose 6-phosphate dehydrogenase are especially sensitive to oxidative stress.



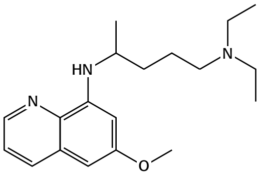
Glutathione (reduced)
(γ-Glutamylcysteinylglycine)

Glucose 6-phosphate dehydrogenase deficiency causes a drug-induced hemolytic anemia

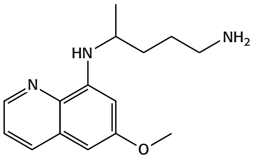
The importance of the pentose phosphate pathway is highlighted by some people's anomalous responses to certain drugs. For instance, pamaquine, the first synthetic antimalarial drug introduced in 1926, was associated with the appearance of severe and mysterious ailments. Most patients tolerated the drug well, but a few developed severe symptoms within a few days after therapy was started. Their urine turned black, jaundice developed, and the hemoglobin content of the blood dropped sharply. In some cases, massive destruction of red blood cells caused death.

This drug-induced *hemolytic anemia* was shown 30 years later to be caused by a *deficiency of glucose 6-phosphate dehydrogenase,* the enzyme catalyzing the first step in the oxidative branch of the pentose phosphate pathway. The result is a dearth of NADPH in all cells, but this deficiency is most acute in red blood cell because they lack mitochondria and have no alternative means of generating reducing power. This defect, which is inherited on the X chromosome, is the most common disease that results from an enzyme malfunction, affecting hundreds of millions of people.

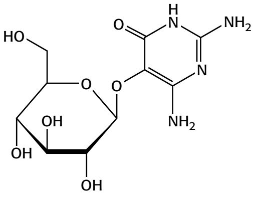
Pamaquine sensitivity is not simply a historical oddity about malaria treatment many decades ago. Primaquine, an antimalarial closely related to pamaquine, is widely used in malaria-infested regions of the world. Vicine, a purine glycoside of fava beans (which are consumed in countries surrounding the Mediterranean), also induces hemolysis. People deficient in glucose 6-phosphate dehydrogenase suffer hemolysis from eating fava beans or inhaling the pollen of the fava flowers, a response called *favism.* How can we explain hemolysis caused by pamaquine, primaquine, and vicine biochemically? These chemicals are oxidative agents that generate peroxides, reactive oxygen species that can damage membranes as well as other biomolecules. Peroxides are usually eliminated by the enzyme *glutathione peroxidase,* which uses reduced glutathione as a reducing agent.



Pamaquine



Primaquine

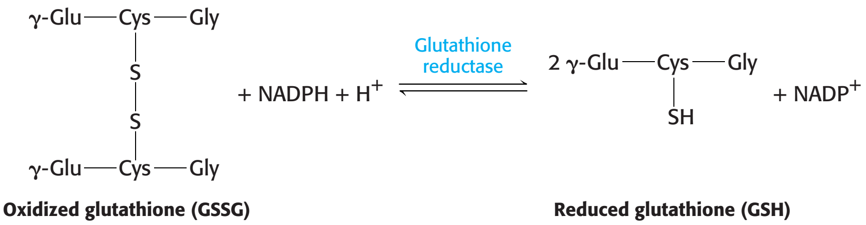


Vicine

Vicia faba. The Mediterranean plant *Vicia faba* is a source of fava beans that contain the purine glycoside vicine. [FLPA/RichardBecker/age footstock.]

$$2\ \mathrm{GSH} + \mathrm{ROOH} \xrightarrow{\text{Glutathione peroxidase}} \mathrm{GSSG} + \mathrm{H_2O} + \mathrm{ROH}$$

The major role of NADPH in red cells is to reduce the disulfide form of glutathione to the sulfhydryl form. The enzyme that catalyzes the regeneration of reduced glutathione is *glutathione reductase*.



Red blood cells with a lowered level of reduced glutathione are more susceptible to hemolysis. In the absence of glucose 6-phosphate dehydrogenase, peroxides continue to damage membranes because no NADPH is being produced to restore reduced glutathione. Thus, the answer to our question is that *glucose 6-phosphate dehydrogenase is required to maintain reduced glutathione levels to protect against oxidative stress*. In the absence of oxidative stress, however, the deficiency is quite benign. The sensitivity to oxidative agents of people having this dehydrogenase deficiency also clearly demonstrates that *atypical reactions to drugs may have a genetic basis*.

Reduced glutathione is also essential for maintaining the normal structure of red blood cells by maintaining the structure of hemoglobin.

The reduced form of glutathione serves as a sulfhydryl buffer that keeps the residues of hemoglobin in the reduced sulfhydryl form. Without adequate levels of reduced glutathione, the hemoglobin sulfhydryl groups can no longer be maintained in the reduced form. Hemoglobin molecules then cross-link with one another to form aggregates called *Heinz bodies* on cell membranes (Figure 20.24). Membranes damaged by Heinz bodies and reactive oxygen species become deformed, and the cell is likely to undergo lysis.

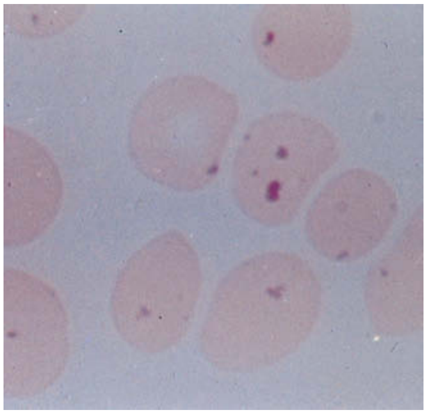


FIGURE 20.24 Red blood cells with Heinz bodies. The light micrograph shows red blood cells obtained from a person deficient in glucose 6-phosphate dehydrogenase. The dark particles, called *Heinz bodies,* inside the cells are clumps of denatured hemoglobin that adhere to the plasma membrane and stain with basic dyes. Red blood cells in such people are highly susceptible to oxidative damage. [Courtesy of Dr. Stanley Schrier.]

A deficiency of glucose 6-phosphate dehydrogenase confers an evolutionary advantage in some circumstances

The incidence of the most common form of glucose 6-phosphate dehydrogenase deficiency, characterized by a 10-fold reduction in enzymatic activity in red blood cells, is 11% among Americans of African heritage. This high frequency suggests that the deficiency may be advantageous under certain environmental conditions. Indeed, *glucose 6-phosphate dehydrogenase deficiency protects against falciparum malaria*. The parasites causing this disease require NADPH for growth. Moreover, infection by the parasites induces oxidative stress in the infected cell. Because the pentose phosphate pathway is compromised, the cell and parasite die from oxidative damage.Thus, glucose 6-phosphate dehydrogenase deficiency is a mechanism of protection against malaria, which accounts for its high frequency in malaria-infested regions of the world. We see here once again the interplay of heredity and environment in the production of disease.

The ability of glucose 6-phosphate dehydrogenase deficiency to protect against malaria does, however, create a public health conundrum. Primaquine is a commonly used and highly effective antimalarial drug. However, indiscriminate use of primaquine causes hemolysis in individuals deficient in glucose 6-phosphate dehydrogenase. A solution to this problem may be in the offing, as recent work shows that an anti-malaria vaccine may be within reach.

SUMMARY

20.1 The Calvin Cycle Synthesizes Hexoses from Carbon Dioxide and Water

ATP and NADPH formed in the light reactions of photosynthesis are used to convert CO_2 into hexoses and other organic compounds. The dark phase of photosynthesis, called the *Calvin cycle,* starts with the reaction of CO_2 and ribulose 1,5-bisphosphate to form two molecules of 3-phosphoglycerate.This reaction is catalyzed by rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase). The steps in the conversion of 3-phosphoglycerate into fructose 6-phosphate and glucose 6-phosphate are like those of gluconeogenesis, except that glyceraldehyde 3-phosphate dehydrogenase in chloroplasts is specific for NADPH rather than NADH. Ribulose 1,5-bisphosphate is regenerated from fructose 6-phosphate, glyceraldehyde 3-phosphate, and dihydroxyacetone phosphate by a complex series of reactions. Several of the steps in the regeneration of ribulose 1,5-bisphosphate are like those of the pentose phosphate pathway. Three molecules of ATP and two molecules of NADPH are consumed for each molecule of CO_2 converted into a hexose. Starch in chloroplasts and sucrose in the cytoplasm are the major carbohydrate stores in plants.

Rubisco also catalyzes a competing oxygenase reaction, which produces phosphoglycolate and 3-phosphoglycerate. The recycling of phosphoglycolate leads to the release of CO_2 and further consumption of O_2 in a process called photorespiration.

20.2 The Activity of the Calvin Cycle Depends on Environmental Conditions

Reduced thioredoxin formed by the light-driven transfer of electrons from ferredoxin activates enzymes of the Calvin cycle by reducing disulfide bridges. The light-induced increase in pH and Mg^{2+} level of the stroma is important in stimulating the carboxylation of ribulose 1,5-bisphosphate by rubisco. Photorespiration is minimized in tropical plants, which have an accessory pathway—the C_4 pathway—for concentrating CO_2 at the site of the Calvin cycle. This pathway enables tropical plants to take advantage of high levels of light and minimize the oxygenation of ribulose 1,5-bisphosphate. Other plants use crassulacean acid metabolism (CAM) to prevent dehydration. In CAM plants, the C_4 pathway is active during the night, when the plant exchanges gases with the air. During the day, gas exchange is eliminated and CO_2 is generated from malate stored in vacuoles.

20.3 The Pentose Phosphate Pathway Generates NADPH and Synthesizes Five-Carbon Sugars

Whereas the Calvin cycle is present only in photosynthetic organisms, the pentose phosphate pathway is present in all organisms. The pentose phosphate pathway consists of two phases: an oxidative phase and a nonoxidative phase. The oxidative phase generates NADPH and ribulose 5-phosphate in the cytoplasm. NADPH is used in reductive biosyntheses. The oxidative phase starts with the dehydrogenation of glucose 6-phosphate to form a lactone, which is hydrolyzed to give 6-phosphogluconate and then oxidatively decarboxylated to yield ribulose 5-phosphate. $NADP^+$ is the electron acceptor in both of these oxidations. The nonoxidative phase begins with the isomerization of ribulose 5-phosphate (a ketose) to ribose 5-phosphate (an aldose). Ribose 5-phosphate is converted into glyceraldehyde 3-phosphate and fructose 6-phosphate by *transketolase* and *transaldolase*. These two enzymes create a reversible link between the pentose phosphate pathway and glycolysis. Xylulose 5-phosphate, sedoheptulose 7-phosphate, and erythrose 4-phosphate are intermediates in these interconversions. When NADPH needs are paramount, fructose 6-phosphate and glyceraldehyde 3-phosphate from the nonoxidative phase are used to synthesize glucose 6-phosphate for metabolism in the oxidative phase. In this way, 12 molecules of NADPH can be generated for each molecule of glucose 6-phosphate that is completely oxidized to CO_2.

20.4 The Metabolism of Glucose 6-Phosphate by the Pentose Phosphate Pathway Is Coordinated with Glycolysis

Only the nonoxidative branch of the pathway is significantly active when much more ribose 5-phosphate than NADPH must be synthesized. Under these conditions, fructose 6-phosphate and glyceraldehyde 3-phosphate (formed by the glycolytic pathway) are converted into ribose 5-phosphate without the formation of NADPH. Alternatively, ribose 5-phosphate formed by the oxidative branch can be converted into pyruvate through fructose 6-phosphate and glyceraldehyde 3-phosphate. In this mode, ATP and NADPH are generated, and five of the six carbons of glucose 6-phosphate emerge in pyruvate. The interplay of the glycolytic and pentose phosphate pathways enables the levels of NADPH, ATP, and building blocks such as ribose 5-phosphate and pyruvate to be continuously adjusted to meet cellular needs.

20.5 Glucose 6-Phosphate Dehydrogenase Plays a Key Role in Protection Against Reactive Oxygen Species

NADPH generated by glucose 6-phosphate dehydrogenase maintains the appropriate levels of reduced glutathione required to combat oxidative stress and ensures the proper reducing environment in the cell. Cells with diminished glucose 6-phosphate dehydrogenase activity are especially sensitive to oxidative stress.

KEY TERMS

Calvin cycle (dark reactions) (p. 589)
autotroph (p. 590)
heterotroph (p. 590)
rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase) (p. 591)
peroxisome (microbody) (p. 593)
photorespiration (p. 594)
hexose monophosphate pool (p. 595)
transketolase (p. 595)
aldolase (p. 595)
starch (p. 597)
sucrose (p. 597)
thioredoxin (p. 599)
C_4 pathway (p. 599)
C_4 plant (p. 601)
C_3 plant (p. 601)
crassulacean acid metabolism (CAM) (p. 601)
pentose phosphate pathway (p. 601)
glucose 6-phosphate dehydrogenase (p. 602)
glutathione (p. 610)

PROBLEMS

1. *A vital cycle*. Why is the Calvin cycle crucial to the functioning of all life forms?

2. *Be nice to plants*. Differentiate between autotrophs and heterotrophs.

3. *Cabalistic reactions?* Why are the reactions of the Calvin cycle sometimes referred to as the dark reactions? Do they take place only at night or are they grim, secret reactions?

4. *Compare and contrast*. Identify the similarities and differences between the Krebs cycle and the Calvin cycle.

5. *Labeling experiments*. When Melvin Calvin performed his initial experiments on carbon fixation, he exposed algae to radioactive carbon dioxide. After 5 seconds, only a single organic compound contained radioactivity but, after 60 seconds, many compounds had incorporated radioactivity. (a) What compound initially contained the radioactivity? (b) What compounds contained radioactivity after 60 seconds?

6. *Three-part harmony*. The Calvin cycle can be thought of as occurring in three parts or stages. Describe the stages.

7. *Not always to the swiftest*. Suggest a reason why rubisco might be the most abundant enzyme in the world.

8. *A requirement*. In an atmosphere devoid of CO_2 but rich in O_2, the oxygenase activity of rubisco disappears. Why?

9. *Reduce locally*. Glyceraldehyde 3-phosphate dehydrogenase in chloroplasts uses NADPH to participate in the synthesis of glucose. In gluconeogenesis in the cytoplasm, the isozyme of the dehydrogenase uses NADH. Why is it advantageous for the chloroplast enzyme to use NADPH?

10. *Total eclipse*. An illuminated suspension of the green alga *Chlorella* is actively carrying out photosynthesis. Suppose that the light is suddenly switched off. How would the levels of 3-phosphoglycerate and ribulose 1,5-bisphosphate change in the next minute?

11. *CO_2 deprivation*. An illuminated suspension of *Chlorella* is actively carrying out photosynthesis in the presence of 1% CO_2. The concentration of CO_2 is abruptly reduced to 0.003%. What effect would this reduction have on the levels of 3-phosphoglycerate and ribulose 1,5-bisphosphate in the next minute?

12. *Salvage operation*. Write a balanced equation for the transamination of glyoxylate to yield glycine.

13. *Dog days of August*. Before the days of pampered lawns, most homeowners practiced horticultural Darwinism. A result was that the lush lawns of early summer would often convert into robust cultures of crabgrass in the dog days of August. Provide a possible biochemical explanation for this transition.

14. *Is it hot in here, or is it just me?* Why is the C_4 pathway valuable for tropical plants?

15. *No free lunch*. Explain why maintaining a high concentration of CO_2 in the bundle-sheath cells of C_4 plants is an example of active transport. How much ATP is required per CO_2 to maintain a high concentration of CO_2 in the bundle-sheath cells of C_4 plants?

16. *Breathing pictures?* What is photorespiration, what is its cause, and why is it believed to be wasteful?

17. *Global warming*. C_3 plants are most common in higher latitudes and become less common at latitudes near the equator. The reverse is true of C_4 plants. How might global warming affect this distribution?

18. *Communication*. What are the light-dependent changes in the stroma that regulate the Calvin cycle?

19. *Match 'em.* Match each term with its description.

(a) Calvin cycle ____
(b) Rubisco ____
(c) Carbamate ____
(d) Starch ____
(e) Sucrose ____
(f) Amylose ____
(g) Amylopectin ____
(h) C_3 plants ____
(i) C_4 plants ____
(j) Stomata ____

1. Catalyzes CO_2 fixation
2. Storage form of carbohydrates
3. α-1,4 linkages only
4. 3-Phosphoglycerate is formed after carbon fixation
5. The dark reactions
6. Includes α-1,6 linkages
7. Required for rubisco activity
8. Carbon fixation results in oxaloacetate formation
9. Allow exchange of gases
10. Transport form of carbohydrates

20. *Linked in.* Describe how the pentose phosphate pathway and glycolysis are linked by transaldolase and transketolase.

21. *Biochemical taxonomy.*

(a) Identify 6-phosphoglucono-δ-lactone. _____

(b) Which reactions produce NADPH? _____

(c) Identify ribulose 5-phosphate. _____

(d) What reaction generates CO_2? _____

(e) Identify 6-phosphogluconate. _____

(f) Which reaction is catalyzed by phosphopentose isomerase? _____

(g) Identify ribose 5-phosphate _____

(h) Which reaction is catalyzed by lactonase? _____

(i) Identify glucose 6-phosphate. _____

(j) Which reaction is catalyzed by 6-phosphogluconate dehydrogenase? _____

(k) Which reaction is catalyzed by glucose 6-phosphate dehydrogenase? _____

23. *Recurring decarboxylations.* Which reaction in the citric acid cycle is most analogous to the oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate? What kind of enzyme-bound intermediate is formed in both reactions?

24. *Synthetic stoichiometries.* What is the stoichiometry of the synthesis of (a) ribose 5-phosphate from glucose 6-phosphate without the concomitant generation of NADPH? (b) NADPH from glucose 6-phosphate without the concomitant formation of pentose sugars?

25. *Offal or awful?* Liver and other organ meats contain large quantities of nucleic acids. In the course of digestion, RNA is hydrolyzed to ribose, among other chemicals. Explain how ribose can be used as a fuel.

26. *A required ATP.* The metabolism of glucose 6-phosphate into ribose 5-phosphate by the joint efforts of the pentose phosphate pathway and glycolysis can be summarized by the following equation.

$$\text{5 Glucose 6-phosphate} + \text{ATP} \longrightarrow \text{6 ribose 5-phosphate} + \text{ADP}$$

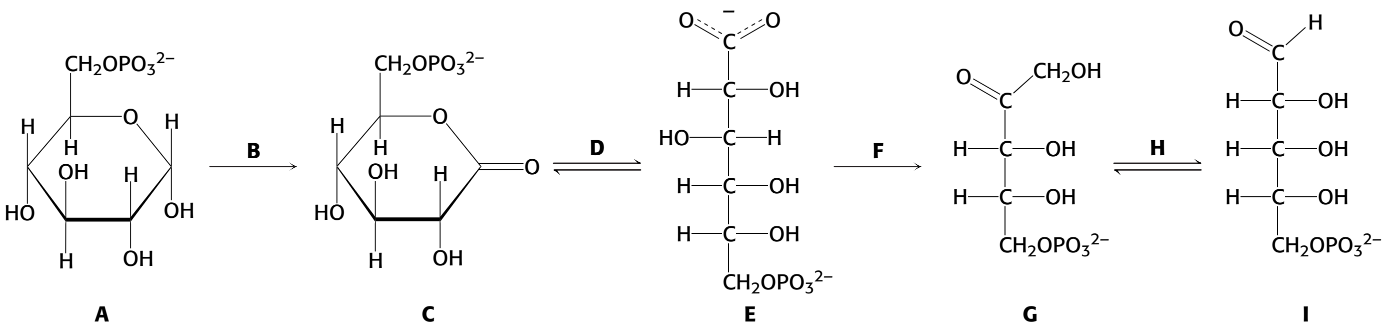
Which reaction requires the ATP?

27. *No respiration.* Glucose is normally completely oxidized to CO_2 in the mitochondria. In what circumstance can glucose be completely oxidized to CO_2 in the cytoplasm?

28. *Watch your diet, doctor.* The noted psychiatrist Hannibal Lecter once remarked to FBI Agent Clarice Starling that he enjoyed liver with some fava beans and a nice Chianti. Why might this diet be dangerous for some people?

29. *No redundancy.* Why do deficiencies in glucose 6-phosphate dehydrogenase frequently present as anemia?

30. *Damage control.* What is the role of glutathione in protection against damage by reactive peroxides? Why is the pentose phosphate pathway crucial to this protection?



22. *Tracing glucose.* Glucose labeled with ^{14}C at C-6 is added to a solution containing the enzymes and cofactors of the oxidative phase of the pentose phosphate pathway. What is the fate of the radioactive label?

31. *Reductive power.* What ratio of NADPH to $NADP^+$ is required to sustain [GSH] = 10 mM and [GSSG] = 1 mM? Use the redox potentials given in Table 18.1.

Mechanism Problems

32. *An alternative approach.* The mechanisms of some aldolases do not include Schiff-base intermediates. Instead, these enzymes require bound metal ions. Propose such a mechanism for the conversion of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate into fructose 1,6-bisphosphate.

33. *A recurring intermediate.* Phosphopentose isomerase interconverts the aldose ribose 5-phosphate and the ketose ribulose 5-phosphate. Propose a mechanism.

Chapter Integration Problems

34. *Catching carbons.* Radioactive-labeling experiments can yield estimates of how much glucose 6-phosphate is metabolized by the pentose phosphate pathway and how much is metabolized by the combined action of glycolysis and the citric acid cycle. Suppose that you have samples of two different tissues as well as two radioactively labeled glucose samples, one with glucose labeled with ^{14}C at C-1 and the other with glucose labeled with ^{14}C at C-6. Design an experiment that would enable you to determine the relative activity of the aerobic metabolism of glucose compared with metabolism by the pentose phosphate pathway.

35. *You do what you can do.* Red blood cells lack mitochondria. These cells process glucose to lactate, but they also generate CO_2. What is the purpose of producing lactate? How can red blood cells generate CO_2 if they lack mitochondria?

36. *Better to focus on one.* Rubisco catalyzes both a carboxylation reaction and a wasteful oxygenase reaction. Below are the kinetic parameters for the two reactions.

$K_M^{CO_2}$ (μM)	$K_M^{O_2}$ (μM)	$k_{cat}^{CO_2}$ (s^{-1})	$k_{cat}^{O_2}$ (s^{-1})
10	500	3	2

(a) Determine the values of $k_{cat}^{CO_2}/K_M^{CO_2}$ and $k_{cat}^{O_2}/K_M^{O_2}$ as $s^{-1}M^{-1}$.

(b) In light of the k_{cat}/K_M values for the two reactions, why does the oxygenation reaction occur?

37. *Photosynthetic efficiency.* Use the following information to estimate the efficiency of photosynthesis.

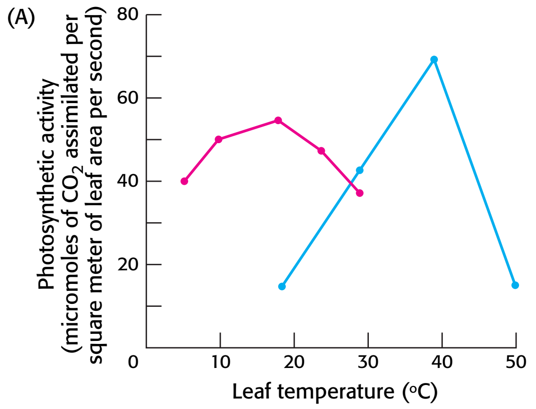
The $\Delta G°'$ for the reduction of CO_2 to the level of hexose is $+477$ kJ mol^{-1} ($+114$ kcal mol^{-1}).
A mole of 600-nm photons has an energy content of 199 kJ (47.6 kcal).

Assume that the proton gradient generated in producing the required NADPH is sufficient to drive the synthesis of the required ATP.

38. *A violation of the First Law?* The complete combustion of glucose to CO_2 and H_2O yields 30 ATP, as shown in Table 18.4. However, the synthesis of glucose requires only 18 ATP. How is it possible that glucose synthesis from CO_2 and H_2O requires only 18 ATP, but combustion to CO_2 and H_2O yields 30 ATP? Is it a violation of the First Law of Thermodynamics, or perhaps a miracle?

Data Interpretation Problem

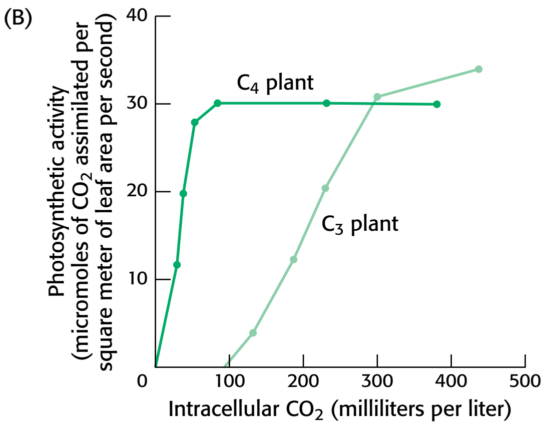
39. *Deciding between 3 and 4.* Graph A shows the photosynthetic activity of two species of plant, one a C_4 plant and the other a C_3 plant, as a function of leaf temperature.



(a) Which data were most likely generated by the C_4 plant and which by the C_3 plant? Explain.

(b) Suggest some possible explanations for why the photosynthetic activity falls at higher temperatures.

Graph B illustrates how the photosynthetic activity of C_3 and C_4 plants varies with CO_2 concentration when temperature (30°C) and light intensity (high) are constant.

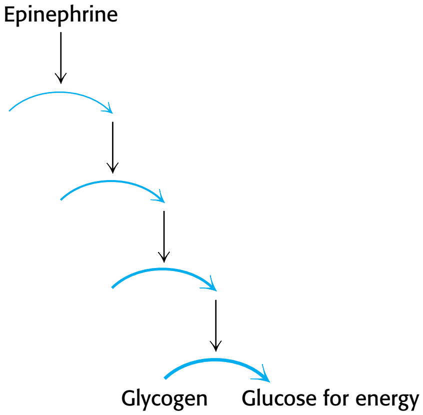


(c) Why can C_4 plants thrive at CO_2 concentrations that do not support the growth of C_3 plants?

(d) Suggest a plausible explanation why C_3 plants continue to increase photosynthetic activity at higher CO_2 concentrations, whereas C_4 plants reach a plateau.

CHAPTER 21

Glycogen Metabolism



Signaling cascades lead to the mobilization of glycogen to produce glucose, an energy source for cyclists. [(Left) Jonathan Devich/Epic Images.]

OUTLINE

21.1 Glycogen Breakdown Requires the Interplay of Several Enzymes

21.2 Phosphorylase Is Regulated by Allosteric Interactions and Reversible Phosphorylation

21.3 Epinephrine and Glucagon Signal the Need for Glycogen Breakdown

21.4 Glycogen Is Synthesized and Degraded by Different Pathways

21.5 Glycogen Breakdown and Synthesis Are Reciprocally Regulated

Glucose is an important fuel and, as we will see, a key precursor for the biosynthesis of many molecules. However, glucose cannot be stored, because high concentrations of glucose disrupt the osmotic balance of the cell, which would cause cell damage or death. How can adequate stores of glucose be maintained without damaging the cell? The solution to this problem is to store glucose as a nonosmotically active polymer called *glycogen*.

Glycogen is a *readily mobilized storage form of glucose.* It is a very large, branched polymer of glucose residues that can be broken down to yield glucose molecules when energy is needed (Figure 21.1). A glycogen molecule has approximately 12 layers of glucose molecules and can be as large as 40 nm, containing approximately 55,000 glucose residues. Most of the glucose residues in glycogen are linked by α-1,4-glycosidic bonds (Figure 21.2). Branches at about every 12 residues are created by α-1,6-glycosidic bonds. Recall that α-glycosidic linkages form open helical polymers, whereas β linkages produce nearly straight strands that form structural fibrils, as in cellulose (Figure 11.14).

Glycogen is not as reduced as fatty acids are and consequently are not as energy rich. Why isn't all excess fuel stored as fatty acids rather than as glycogen? The controlled release of glucose from glycogen maintains blood-glucose concentration between meals. The circulating blood keeps the brain supplied with glucose, which is virtually the only fuel used by the brain, except during prolonged starvation (Chapter 27). Moreover, the readily mobilized glucose from glycogen is a good source of energy for sudden, strenuous activity. Unlike fatty acids, the released glucose can be metabolized in the absence of oxygen and can thus supply energy for anaerobic activity (Section 16.1).

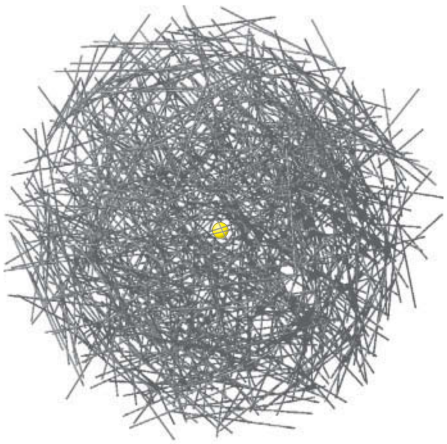


FIGURE 21.1 Glycogen. At the core of the glycogen molecule is the protein glycogenin (yellow). Each line represents glucose molecules joined by an α-1,4-glycosidic linkage. The nonreducing ends of the glycogen molecule form the surface of the glycogen granule. Degradation takes place at this surface. [Information from R. Melendez et al. *Biophys. J.* 77:1327–1332, 1999.]

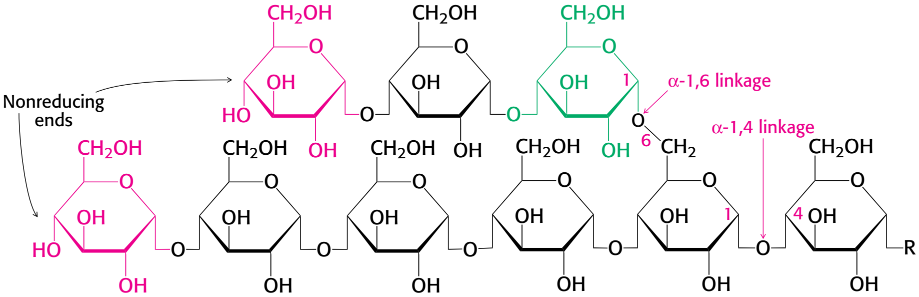

Nonreducing ends
α-1,6 linkage
α-1,4 linkage
CH₂OH
OH
HO
O
CH₂
1
6
4
R

FIGURE 21.2 Glycogen structure. In this structure of two outer branches of a glycogen molecule, the residues at the nonreducing ends are shown in red and the residue that starts a branch is shown in green. The rest of the glycogen molecule is represented by R.

Glycogen is present in bacteria, archae, and eukaryotes. Recall that plants store glucose as starch, a similar chemical. Thus, storing energy as glucose polymers is common to all forms of life. In humans, most tissues have some glycogen, although the two major sites of glycogen storage are the liver and skeletal muscle. The concentration of glycogen is higher in the liver than in muscle (10% versus 2% by weight), but more glycogen is stored in skeletal muscle overall because of muscle's much greater mass. Glycogen is present in the cytoplasm, with the molecule appearing as granules (Figure 21.3). In the liver, glycogen synthesis and degradation are regulated to maintain blood-glucose levels as required to meet the needs of the organism as a whole. In contrast, in muscle, these processes are regulated to meet the energy needs of the muscle itself.

Glycogen metabolism is the regulated release and storage of glucose

Glycogen degradation and synthesis are simple biochemical processes. Glycogen degradation consists of three steps: (1) the release of glucose 1-phosphate from glycogen, (2) the remodeling of the glycogen substrate to permit further degradation, and (3) the conversion of glucose 1-phosphate into glucose 6-phosphate for further metabolism. The glucose 6-phosphate derived from the breakdown of glycogen has three possible fates (Figure 21.4): (1) It can be metabolized by glycolysis, (2) it can be converted into free

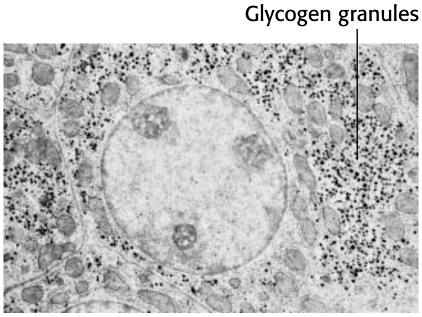

Glycogen granules

FIGURE 21.3 Electron micrograph of a liver cell. The dark particles in the cytoplasm are glycogen granules. [Courtesy of Dr. George Palade.]

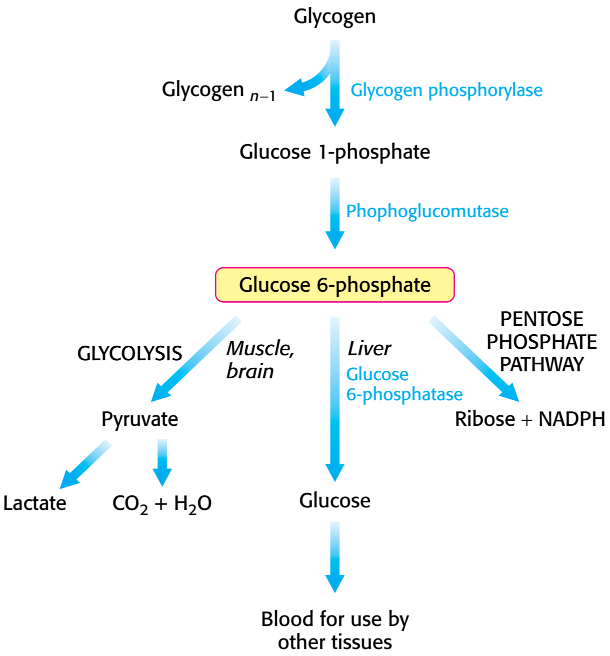

Glycogen
Glycogen $_{n-1}$
Glycogen phosphorylase
Glucose 1-phosphate
Phophoglucomutase
Glucose 6-phosphate
GLYCOLYSIS
Muscle, brain
Liver
Glucose 6-phosphatase
PENTOSE PHOSPHATE PATHWAY
Pyruvate
Ribose + NADPH
Lactate
CO₂ + H₂O
Glucose
Blood for use by other tissues

FIGURE 21.4 Fates of glucose 6-phosphate. Glucose 6-phosphate derived from glycogen can (1) be used as a fuel for anaerobic or aerobic metabolism as in, for instance, muscle; (2) be converted into free glucose in the liver and subsequently released into the blood; (3) be processed by the pentose phosphate pathway to generate NADPH or ribose in a variety of tissues.

glucose for release into the bloodstream, and (3) it can be processed by the pentose phosphate pathway to yield NADPH and ribose derivatives. The conversion of glycogen into free glucose takes place mainly in the liver.

Glycogen synthesis, which takes place when glucose is abundant, requires an activated form of glucose, uridine diphosphate glucose (UDP-glucose), formed by the reaction of UTP and glucose 1-phosphate. As is the case for glycogen degradation, the glycogen molecule must be remodeled for continued synthesis.

The regulation of glycogen degradation and synthesis is complex and is facilitated by the fact that all of the enzymes involved in glycogen metabolism and its regulation are associated with the glycogen particle. Several enzymes taking part in glycogen metabolism allosterically respond to metabolites that signal the energy needs of the cell. *Through these allosteric responses, enzyme activity is adjusted to meet the needs of the cell.* In addition, hormones may initiate signal cascades that lead to the reversible phosphorylation of enzymes, which alters their catalytic rates. *Regulation by hormones adjusts glycogen metabolism to meet the needs of the entire organism.*

21.1 Glycogen Breakdown Requires the Interplay of Several Enzymes

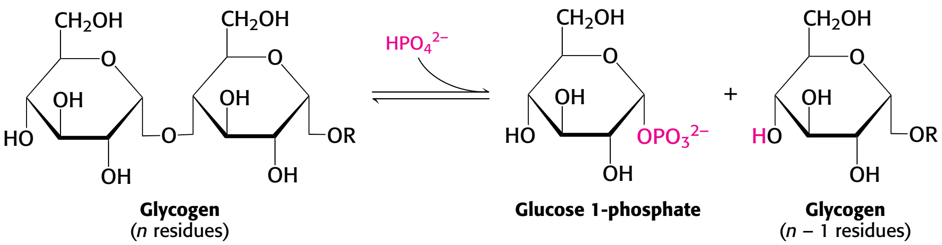
The efficient breakdown of glycogen to provide glucose 6-phosphate for further metabolism requires four enzyme activities: one to degrade glycogen, two to remodel glycogen so that it can remain a substrate for degradation, and one to convert the product of glycogen breakdown into a form suitable for further metabolism. We will examine each of these activities in turn.

Phosphorylase catalyzes the phosphorolytic cleavage of glycogen to release glucose 1-phosphate

Glycogen phosphorylase, the key enzyme in glycogen breakdown, cleaves its substrate by the addition of orthophosphate (P_i) to yield *glucose 1-phosphate.* The cleavage of a bond by the addition of orthophosphate is referred to as *phosphorolysis*.

$$\underset{(n \text{ residues})}{\text{Glycogen}} + P_i \rightleftharpoons \text{glucose 1-phosphate} + \underset{(n-1 \text{ residues})}{\text{glycogen}}$$

Phosphorylase catalyzes the sequential removal of glucosyl residues from the nonreducing ends of the glycogen molecule (the ends with a free OH group on carbon 4). Orthophosphate splits the glycosidic linkage between C-1 of the terminal residue and C-4 of the adjacent one. Specifically, it cleaves the bond between the C-1 carbon atom and the glycosidic oxygen atom, and the α configuration at C-1 is retained.



Glucose 1-phosphate released from glycogen can be readily converted into glucose 6-phosphate, an important metabolic intermediate, by the enzyme phosphoglucomutase.

The reaction catalyzed by phosphorylase is readily reversible in vitro. At pH 6.8, the equilibrium ratio of orthophosphate to glucose 1-phosphate is 3.6. The value of $\Delta G^{\circ\prime}$ for this reaction is small because a glycosidic bond is replaced by a phosphoryl ester bond that has a nearly equal transfer potential. However, phosphorolysis proceeds far in the direction of glycogen breakdown in vivo because the $[P_i]$/[glucose 1-phosphate] ratio is usually greater than 100, substantially favoring phosphorolysis. We see here an example of how the cell can alter the free-energy change to favor a reaction's occurrence by altering the ratio of substrate and product.

The phosphorolytic cleavage of glycogen is energetically advantageous because the released sugar is already phosphorylated. In contrast, a hydrolytic cleavage would yield glucose, which would then have to be phosphorylated at the expense of a molecule of ATP to enter the glycolytic pathway. An additional advantage of phosphorolytic cleavage for muscle cells is that no transporters exist for glucose 1-phosphate, which is negatively charged under physiological conditions, and so it cannot be transported or diffuse out of the cell.

Mechanism: Pyridoxal phosphate participates in the phosphorolytic cleavage of glycogen

The special challenge faced by phosphorylase is to cleave glycogen phosphorolytically rather than hydrolytically to save the ATP required to phosphorylate free glucose. Thus, water must be excluded from the active site. Phosphorylase is a dimer of two identical 97-kDa subunits. Each subunit is compactly folded into an *amino-terminal domain* (480 residues) containing a *glycogen-binding site* and a *carboxyl-terminal domain* (360 residues; Figure 21.5). The catalytic site in each subunit is located in a deep crevice formed by residues from both domains. Substrates bind synergistically, causing the crevice to narrow, thereby excluding water. What is the mechanistic basis of the phosphorolytic cleavage of glycogen?

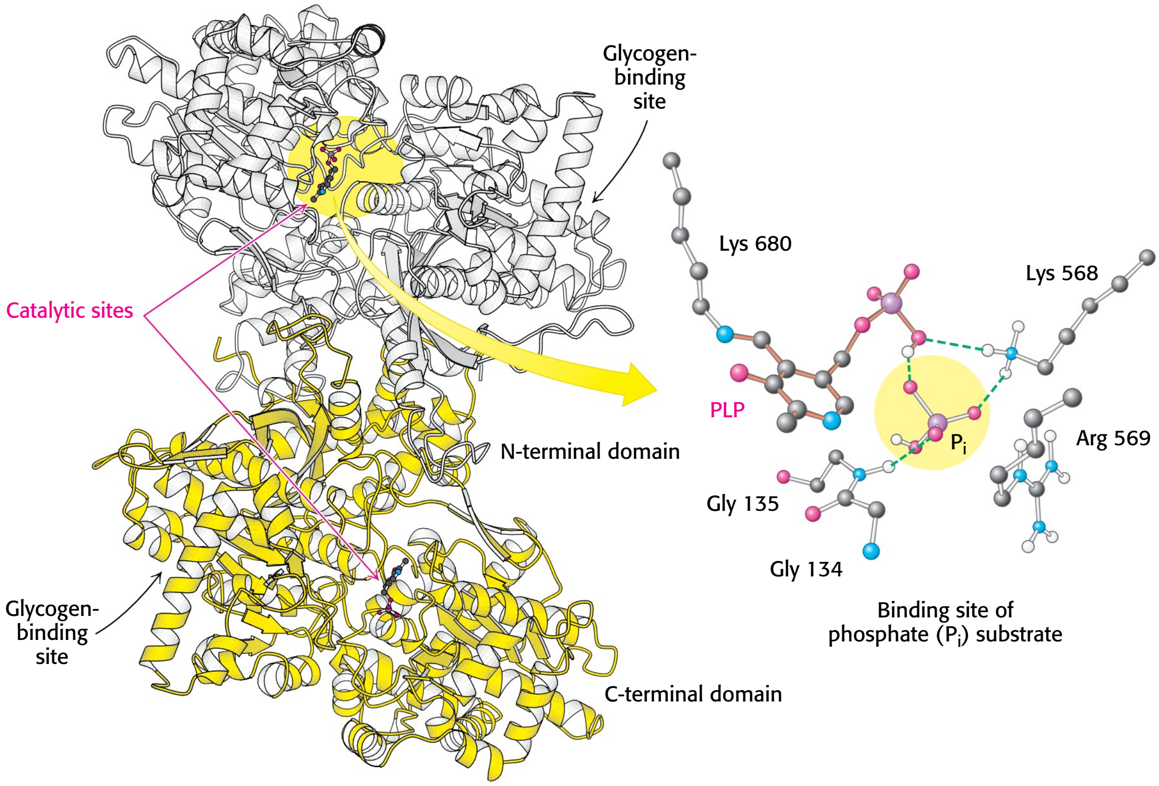


FIGURE 21.5 Structure of glycogen phosphorylase. This enzyme forms a homodimer: one subunit is shown in white and the other in yellow. Each catalytic site includes a pyridoxal phosphate (PLP) group, linked to lysine 680 of the enzyme. The binding site for the phosphate (P_i) substrate is shown. *Notice* that the catalytic site lies between the C-terminal domain and the glycogen-binding site. A narrow crevice, which binds four or five glucose units of glycogen, connects the two sites. The separation of the sites allows the catalytic site to phosphorolyze several glucose units before the enzyme must rebind the glycogen substrate. [Drawn from 1NOI.pdb.]

Several clues suggest a mechanism. First, both the glycogen substrate and the glucose 1-phosphate product have an α configuration at C-1. A direct attack by phosphate on C-1 of a sugar would invert the configuration at this carbon atom because the reaction would proceed through a pentacovalent transition state. The fact that the glucose 1-phosphate formed has an α rather than a β configuration suggests that an even number of steps (most simply, two) is required. A likely explanation for these results is that a *carbonium ion intermediate* is formed from the glucose residue.

A second clue to the catalytic mechanism of phosphorylase is its requirement for the coenzyme *pyridoxal phosphate* (PLP), a derivative of pyridoxine (vitamin B_6, Section 15.4). The aldehyde group of this coenzyme forms a Schiff-base linkage with a specific lysine side chain of the enzyme (Figure 21.6). Structural studies indicate that the reacting orthophosphate group takes a position between the 5′-phosphate group of PLP and the glycogen substrate (Figure 21.7). *The 5′-phosphate group of PLP acts in tandem with orthophosphate by serving as a proton donor and then as a proton acceptor (i.e., as a general acid–base catalyst).* Orthophosphate (in the HPO_4^{2-} form) donates a proton to the oxygen atom attached to carbon 4 of the departing glycogen chain and simultaneously acquires a proton from PLP. The carbonium ion intermediate formed in this step is then attacked by orthophosphate to form α-glucose 1-phosphate, with the concomitant return of a proton to pyridoxal phosphate. The special role of pyridoxal phosphate in the reaction is necessary because water is excluded from the active site.

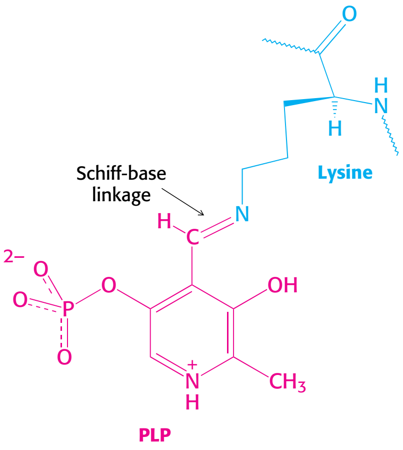


FIGURE 21.6 PLP–Schiff-base linkage. A pyridoxal phosphate (PLP) group (red) forms a Schiff base with a lysine residue (blue) at the active site of phosphorylase, where it functions as a general acid–base catalyst.

A Schiff base, also called an *imine,* is a compound containing a carbon–nitrogen double bond with the nitrogen atom bonded to an organic compound, not a hydrogen atom. A Schiff base is formed by the reaction of a primary amine with an aldehyde or ketone.

The glycogen-binding site is 30 Å away from the catalytic site (Figure 21.5), but it is connected to the catalytic site by a narrow crevice able to accommodate four or five glucose units. The large separation between the binding site and the catalytic site enables the enzyme to phosphorolyze many residues without having to dissociate and reassociate after each catalytic cycle. An enzyme that can catalyze many reactions without having to dissociate and reassociate after each catalytic step is said to be *processive*—a property of enzymes that synthesize and degrade large polymers. We will see such enzymes again when we consider DNA and RNA synthesis.

A debranching enzyme also is needed for the breakdown of glycogen

Glycogen phosphorylase acting alone degrades glycogen to a limited extent. The enzyme can break α-1,4-glycosidic bonds on glycogen branches but soon encounters an obstacle. The α-1,6-glycosidic bonds at the branch points are not susceptible to cleavage by phosphorylase. Indeed, phosphorylase stops cleaving α-1,4 linkages when it reaches a terminal residue four residues away from a branch point. Because about 1 in 12 residues is branched, cleavage by the phosphorylase alone would come to a halt after the release of eight glucose molecules per branch.

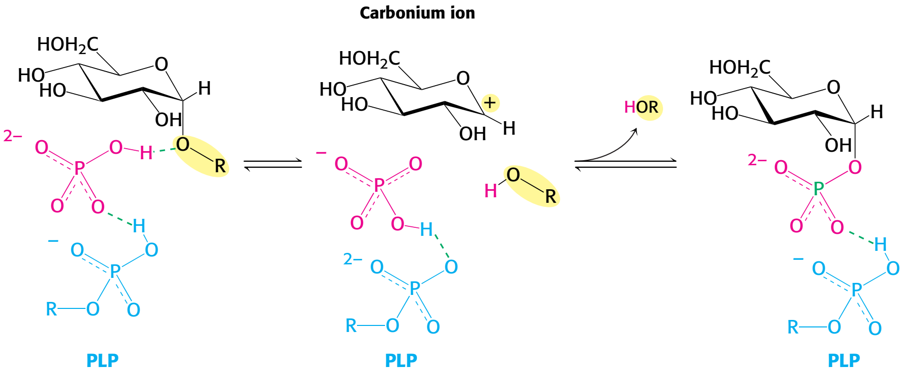
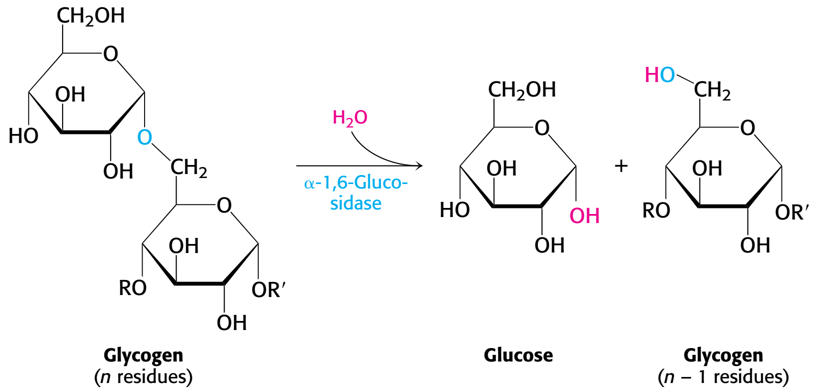


FIGURE 21.7 Phosphorylase mechanism. A bound HPO_4^{2-} group (red) favors the cleavage of the glycosidic bond by donating a proton to the C-4 oxygen of the departing glycosyl group (black). This reaction results in the formation of a carbonium ion and is favored by the transfer of a proton from the protonated phosphate group of the bound pyridoxal phosphate (PLP) group (blue). The carbonium ion and the orthophosphate combine to form glucose 1-phosphate.

How can the remainder of the glycogen molecule be mobilized for use as a fuel? Two additional enzymes, a *transferase* and *α-1,6-glucosidase*, remodel the glycogen for continued degradation by the phosphorylase (Figure 21.8). *The transferase shifts a block of three glucosyl residues from one outer branch to another.* This transfer exposes a single glucose residue joined by an α-1,6-glycosidic linkage. α-1,6-Glucosidase, also known as the debranching enzyme, hydrolyzes the α-1,6-glycosidic bond.



A free glucose molecule is released and then phosphorylated by the glycolytic enzyme hexokinase. Thus, the transferase and α-1,6-glucosidase convert the branched structure into a linear one, which paves the way for further cleavage by phosphorylase. In eukaryotes, the transferase and the α-1,6-glucosidase activities are present in a single 160-kDa polypeptide chain, providing yet another example of a bifunctional enzyme (Figure 16.29).

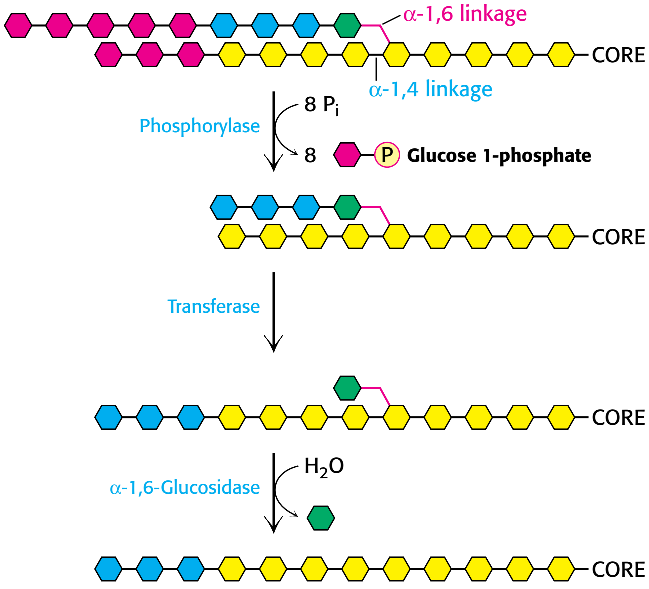


FIGURE 21.8 Glycogen remodeling. First, α-1,4-glycosidic bonds on each branch are cleaved by phosphorylase, leaving four residues along each branch. The transferase shifts a block of three glucosyl residues from one outer branch to the other. In this reaction, the α-1,4-glycosidic link between the blue and the green residues is broken and a new α-1,4 link between the blue and the yellow residues is formed. The green residue is then removed by α-1,6-glucosidase, leaving a linear chain with all α-1,4 linkages, suitable for further cleavage by phosphorylase.

Phosphoglucomutase converts glucose 1-phosphate into glucose 6-phosphate

Glucose 1-phosphate formed in the phosphorolytic cleavage of glycogen must be converted into glucose 6-phosphate to enter the metabolic mainstream. This shift of a phosphoryl group is catalyzed by *phosphoglucomutase.* Recall that this enzyme is also used in galactose metabolism (Section 16.1). To effect this shift, the enzyme exchanges a phosphoryl group with the substrate (Figure 21.9). The catalytic site of an active mutase molecule contains a phosphorylated serine residue. The phosphoryl group is transferred from the serine residue to the C-6 hydroxyl group of glucose 1-phosphate to form glucose 1,6-bisphosphate. The C-1 phosphoryl group of this intermediate is then shuttled to the same serine residue, resulting in the formation of glucose 6-phosphate and the regeneration of the phosphoenzyme.

These reactions are like those of *phosphoglycerate mutase,* a glycolytic enzyme (Section 16.1). The role of glucose 1,6-bisphosphate in the interconversion of the phosphoglucoses is like that of 2,3-bisphosphoglycerate (2,3-BPG) in the interconversion of 2-phosphoglycerate and 3-phosphoglycerate in glycolysis. A phosphoenzyme intermediate participates in both reactions.

The liver contains glucose 6-phosphatase, a hydrolytic enzyme absent from muscle

A major function of the liver is to maintain a nearly constant level of glucose in the blood. The liver releases glucose into the blood between meals and

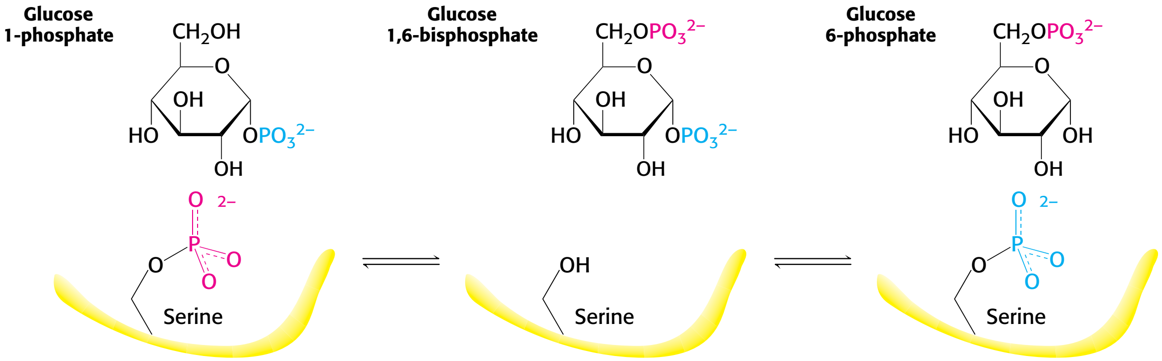


FIGURE 21.9 Reaction catalyzed by phosphoglucomutase. A phosphoryl group is transferred from the enzyme to the substrate, and a different phosphoryl group is transferred back to restore the enzyme to its initial state.

during muscular activity. The released glucose is taken up primarily by the brain, skeletal muscle, and red blood cells. In contrast with unmodified glucose, however, the phosphorylated glucose produced by glycogen breakdown is not transported out of cells. The liver contains a hydrolytic enzyme, *glucose 6-phosphatase,* that enables glucose to leave that organ. This enzyme cleaves the phosphoryl group to form free glucose and orthophosphate. This glucose 6-phosphatase is the same enzyme that releases free glucose at the conclusion of gluconeogenesis. It is located on the lumenal side of the smooth endoplasmic reticulum membrane. Recall that glucose 6-phosphate is transported into the endoplasmic reticulum; glucose and orthophosphate formed by hydrolysis are then shuttled back into the cytoplasm (Section 16.1).

$$\text{Glucose 6-phosphate} + H_2O \longrightarrow \text{glucose} + P_i$$

Glucose 6-phosphatase is absent from most other tissues. Muscle tissues retain glucose 6-phosphate for the generation of ATP. In contrast, glucose is not a major fuel for the liver.

21.2 Phosphorylase Is Regulated by Allosteric Interactions and Reversible Phosphorylation

Glycogen degradation is precisely controlled by multiple interlocking mechanisms. The focus of this control is the enzyme glycogen phosphorylase. *Phosphorylase is regulated by several allosteric effectors that signal the energy state of the cell as well as by reversible phosphorylation, which is responsive to hormones such as insulin, epinephrine, and glucagon.* We will examine the differences in the control of two isozymic forms of glycogen phosphorylase: one specific to liver and one specific to skeletal muscle. These differences are due to the fact that *the liver maintains glucose homeostasis of the organism as a whole, whereas the muscle uses glucose to produce energy for itself.*

Liver phosphorylase produces glucose for use by other tissues

The dimeric phosphorylase exists in two interconvertible forms: a *usually active* phosphorylase *a* and a *usually inactive* phosphorylase *b* (Figure 21.10). Each of these two forms exists in equilibrium between an active relaxed (R) state and a much less active tense (T) state, but the equilibrium for phosphorylase *a* favors the active R state, whereas the equilibrium for phosphorylase *b* favors the less-active T state (Figure 21.11). The role of glycogen degradation in the liver is to form glucose for export to other tissues when the blood-glucose level is low. Consequently, we can think of the

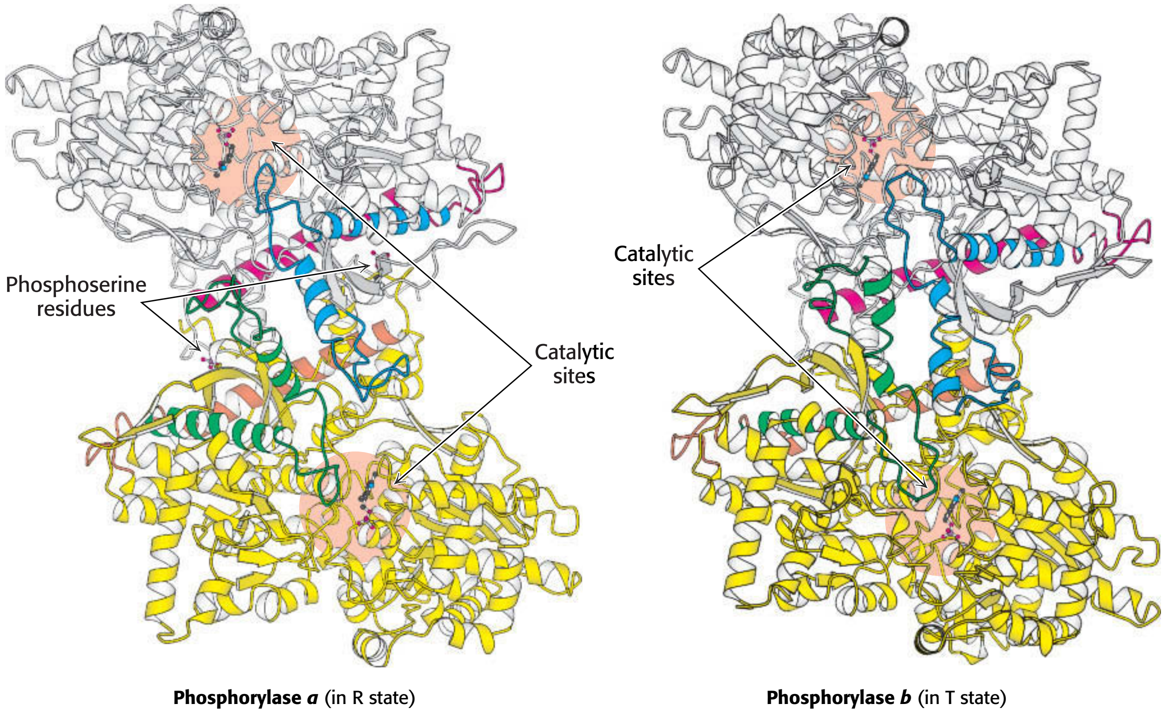

Phosphoserine residues
Catalytic sites
Catalytic sites
Phosphorylase **a** (in R state)
Phosphorylase **b** (in T state)

FIGURE 21.10 Structures of phosphorylase *a* and phosphorylase *b*. Phosphorylase *a* is phosphorylated on serine 14 of each subunit. This modification favors the structure of the more-active R state. One subunit is shown in white, with helices and loops important for regulation shown in blue and red. The other subunit is shown in yellow, with the regulatory structures shown in orange and green. Phosphorylase *b* is not phosphorylated and exists predominantly in the T state. *Notice* that the catalytic sites are partly occluded in the T state. [Drawn from 1GPA.pdb and 1NOJ.pdb.]

default state of liver phosphorylase as being the *a* form: glucose is to be generated unless the enzyme is signaled otherwise. The liver phosphorylase *a* form thus exhibits the most responsive R ↔ T transition (Figure 21.12). The binding of glucose to the active site shifts the *a* form from the active R state to the less-active T state. In essence, the enzyme reverts to the low-activity T state only when it detects the presence of sufficient glucose. If glucose is present in the diet, there is no need to degrade glycogen.

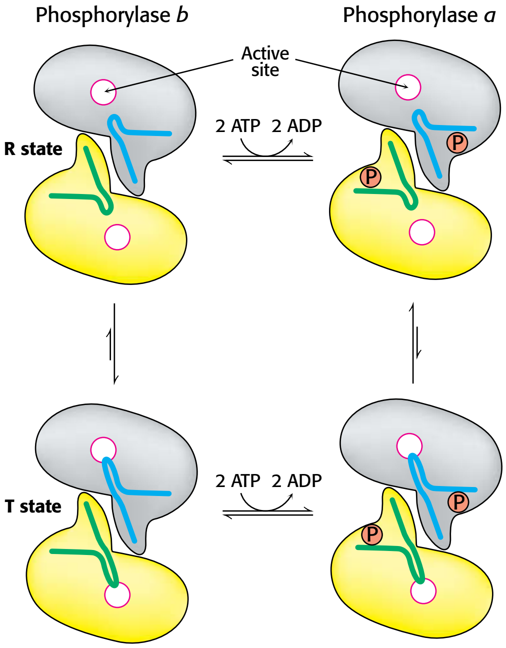

Phosphorylase *b*
Phosphorylase *a*
Active site
R state
2 ATP
2 ADP
P
P
T state
2 ATP
2 ADP
P
P

FIGURE 21.11 Phosphorylase regulation. Both phosphorylase *b* and phosphorylase *a* exist as equilibria between an active R state and a less-active T state. Phosphorylase *b* is usually inactive because the equilibrium favors the T state. Phosphorylase *a* is usually active because the equilibrium favors the R state. Regulatory structures are shown in blue and green.

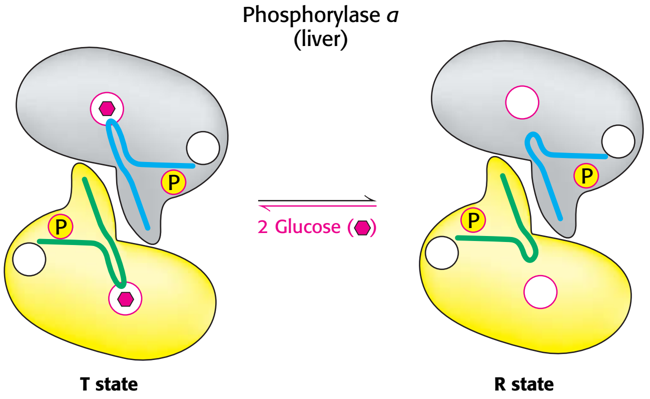

Phosphorylase *a* (liver)
P
P
2 Glucose (⬢)
P
P
T state
R state

FIGURE 21.12 Allosteric regulation of liver phosphorylase. The binding of glucose to phosphorylase *a* shifts the equilibrium to the T state and inactivates the enzyme. Thus, glycogen is not mobilized when glucose is already abundant.

Muscle phosphorylase is regulated by the intracellular energy charge

In contrast to the liver isozyme, the default state of muscle phosphorylase is the *b* form, owing to the fact that, for muscle, phosphorylase must be active primarily during muscle contraction. Muscle phosphorylase *b* is activated by the presence of high concentrations of AMP, which binds to a nucleotide-binding site and stabilizes the conformation of phosphorylase *b* in the active R state (Figure 21.13). Thus, when a muscle contracts and ATP is converted into AMP by the sequential action of myosin (Section 9.4) and adenylate kinase (Section 16.2), the phosphorylase is signaled to degrade glycogen. ATP acts as a negative allosteric effector by competing with AMP. Thus, *the transition of phosphorylase* b *between the active R state and the less-active T state is controlled by the energy charge of the muscle cell.* Glucose 6-phosphate also binds at the same site as ATP and stabilizes the less-active state of phosphorylase *b,* an example of feedback inhibition. Under most physiological conditions, *phosphorylase* b *is inactive because of the inhibitory effects of ATP and glucose 6-phosphate.* In contrast, *phosphorylase* a *is fully active,* regardless of the levels of AMP, ATP, and glucose 6-phosphate. In resting muscle, nearly all the enzyme is in the inactive *b* form.

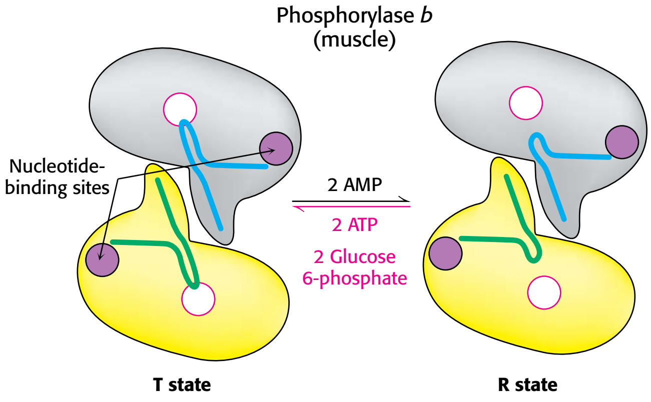


FIGURE 21.13 Allosteric regulation of muscle phosphorylase. A low energy charge, represented by high concentrations of AMP, favors the transition to the R state. ATP and glucose 6-phosphate stabilize the T state.

Unlike the enzyme in muscle, the liver phosphorylase is insensitive to regulation by AMP because the liver does not undergo the dramatic changes in energy charge seen in a contracting muscle. We see here a clear example of the use of isozymes to establish the tissue-specific biochemical properties of the liver and muscle. In human beings, liver phosphorylase and muscle phosphorylase are approximately 90% identical in amino acid sequence, yet the 10% difference results in subtle but important shifts in the stability of various forms of the enzyme.

Biochemical characteristics of muscle fiber types differ

Not only do the biochemical needs of liver and muscle differ with respect to glycogen metabolism, but the biochemical needs of different muscle fiber types also vary. Skeletal muscle consists of three different fiber types: type I, or slow-twitch muscle; type IIb (also called type IIx), or fast-twitch fibers; and type IIa fibers that have properties intermediate between the other two fiber types. Type I fibers rely predominantly on cellular respiration to derive energy. These fibers are powered by fatty acid degradation and are rich in mitochondria, the site of fatty acid degradation and the citric acid cycle. As we will see in Chapter 22, fatty acids are an excellent energy storage form, but generating ATP from fatty acids is slower than from glycogen. Glycogen is not an important fuel for type I fibers, and consequently the amount of glycogen phosphorylase is low. Type I fibers power endurance activities. Type IIb fibers use glycogen as their main fuel. Consequently, the glycogen and glycogen phosphorylase are abundant. These fibers are also rich in glycolytic enzymes needed to process glucose quickly in the absence of oxygen and poor in mitochondria. Type IIb fibers power burst activities such as sprinting and weight lifting. No amount of training can interconvert type I fibers and type IIb fibers. However, there is some evidence that type IIa fibers are "trainable"; that is, endurance training enhances the oxidative capacity of type IIa fibers while burst activity training enhances the glycolytic capacity. Table 21.1 shows the biochemical profile of the fiber types.

TABLE 21.1 Biochemical characteristics of muscle fiber types

Characteristic	Type I	Type IIa	Type IIb
Fatigue resistance	High	Intermediate	Low
Mitochondrial density	High	Intermediate	Low
Metabolic type	Oxidative	Oxidative/glycolytic	Glycolytic
Myoglobin content	High	Intermediate	Low
Glycogen content	Low	Intermediate	High
Triacylglycerol content	High	Intermediate	Low
Glycogen phosphorylase activity	Low	Intermediate	High
Phosphofructokinase activity	Low	Intermediate	High
Citrate synthase activity	High	Intermediate	Low

Phosphorylation promotes the conversion of phosphorylase *b* to phosphorylase *a*

In both liver and muscle, phosphorylase *b* is converted into phosphorylase *a* by the phosphorylation of a single serine residue (serine 14) in each subunit. This conversion is initiated by hormones. Low blood levels of glucose lead to the secretion of the hormone glucagon. Fear or the excitement of exercise will cause levels of the hormone epinephrine to increase. The rise in glucagon and epinephrine levels results in phosphorylation of the enzyme to the phosphorylase *a* form in liver and muscle respectively. The regulatory enzyme *phosphorylase kinase* catalyzes this covalent modification.

Comparison of the structures of phosphorylase *a* in the R state and phosphorylase *b* in the T state reveals that subtle structural changes at the subunit interfaces are transmitted to the active sites (Figure 21.10). The transition from the T state (the prevalent state of phosphorylase *b*) to the R state (the prevalent state of phosphorylase *a*) entails a 10-degree rotation around the twofold axis of the dimer. Most important, this transition is associated with structural changes in α helices that move a loop out of the active site of each subunit. Thus, the T state is less active because the catalytic site is partly blocked. In the R state, the catalytic site is more accessible and a binding site for orthophosphate is well organized.

Phosphorylase kinase is activated by phosphorylation and calcium ions

Phosphorylase kinase activates phosphorylase *b* by attaching a phosphoryl group. The subunit composition of phosphorylase kinase in skeletal muscle is $(\alpha\beta\gamma\delta)_4$, and the mass of this very large protein is 1300 kDa. The enzyme consists of two $(\alpha\beta\gamma\delta)_2$ lobes that are joined by a β_4 bridge that is the core of the enzyme and serves as a scaffold for the remaining subunits. The γ subunit contains the active site, while all of the remaining subunits ($\approx$90% by mass) play regulatory roles. The δ subunit is the calcium binding protein *calmodulin*, a calcium sensor that stimulates many enzymes in eukaryotes (Section 14.1). The α and β subunits are targets of protein kinase A. The β subunit is phosphorylated first, followed by the phosphorylation of the α subunit.

Activation of phosphorylase kinase is initiated when Ca^{2+} binds to the δ subunit. This mode of activation of the kinase is especially noteworthy in muscle, where contraction is triggered by the release of Ca^{2+} from the sarcoplasmic reticulum (Figure 21.14). Maximal activation is achieved with the phosphorylation of the β and α subunits of the Ca^{2+}-bound kinase. The stimulation of phosphorylase kinase is one step in a signal-transduction cascade initiated by signal molecules such as glucagon and epinephrine.

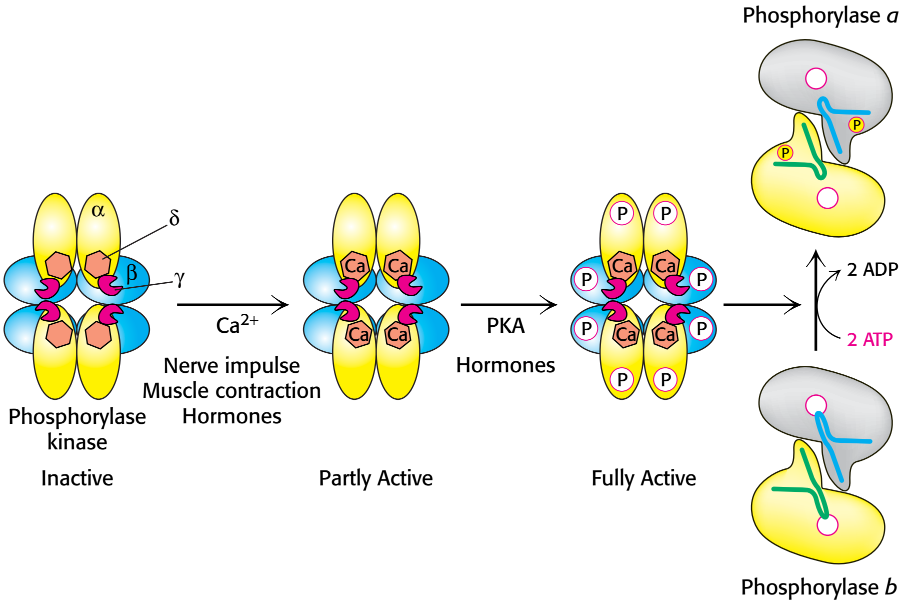


FIGURE 21.14 Activation of phosphorylase kinase. Phosphorylase kinase, an $(\alpha\beta\gamma\delta)_4$ assembly, is partly activated by Ca^{2+} binding to the δ subunit. Activation is maximal when the β and α subunits are phosphorylated in response to hormonal signals. When active, the enzyme converts phosphorylase *b* into phosphorylase *a*.

21.3 Epinephrine and Glucagon Signal the Need for Glycogen Breakdown

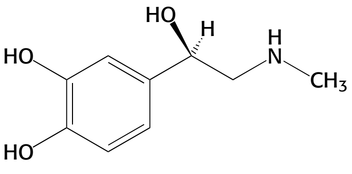
Protein kinase A activates phosphorylase kinase, which in turn activates glycogen phosphorylase. What activates protein kinase A? What is the signal that ultimately triggers an increase in glycogen breakdown?

G proteins transmit the signal for the initiation of glycogen breakdown

Several hormones greatly affect glycogen metabolism. Glucagon and epinephrine trigger the breakdown of glycogen. Muscular activity or its anticipation leads to the release of *epinephrine (adrenaline),* a catecholamine derived from tyrosine, from the adrenal medulla. Epinephrine markedly stimulates glycogen breakdown in muscle and, to a lesser extent, in the liver. The liver is more responsive to *glucagon,* a polypeptide hormone secreted by the α cells of the pancreas when the blood-sugar level is low. Physiologically, glucagon signifies the starved state (Figure 21.15).

How do hormones trigger the breakdown of glycogen? They initiate a cyclic AMP signal-transduction cascade, already discussed in Section 14.1 (Figure 21.16).

1. The signal molecules epinephrine and glucagon bind to specific seven-transmembrane (7TM) receptors in the plasma membranes of target cells (Section 14.1). Epinephrine binds to the β-adrenergic receptor in muscle, whereas glucagon binds to the glucagon receptor in the liver. These binding events activate the G_s protein. *A specific external signal has been transmitted into the cell through structural changes,* first in the receptor and then in the G protein.



Epinephrine

5 10
$^{+}H_3N$–His–Ser–Glu–Gly–Thr–Phe–Thr–Ser–Asp–Tyr–
15 20
–Ser–Lys–Tyr–Leu–Asp–Ser–Arg–Arg–Ala–Gln–
25 29
–Asp–Phe–Val–Gln–Trp–Leu–Met–Asn–Thr–COO^{-}

Glucagon

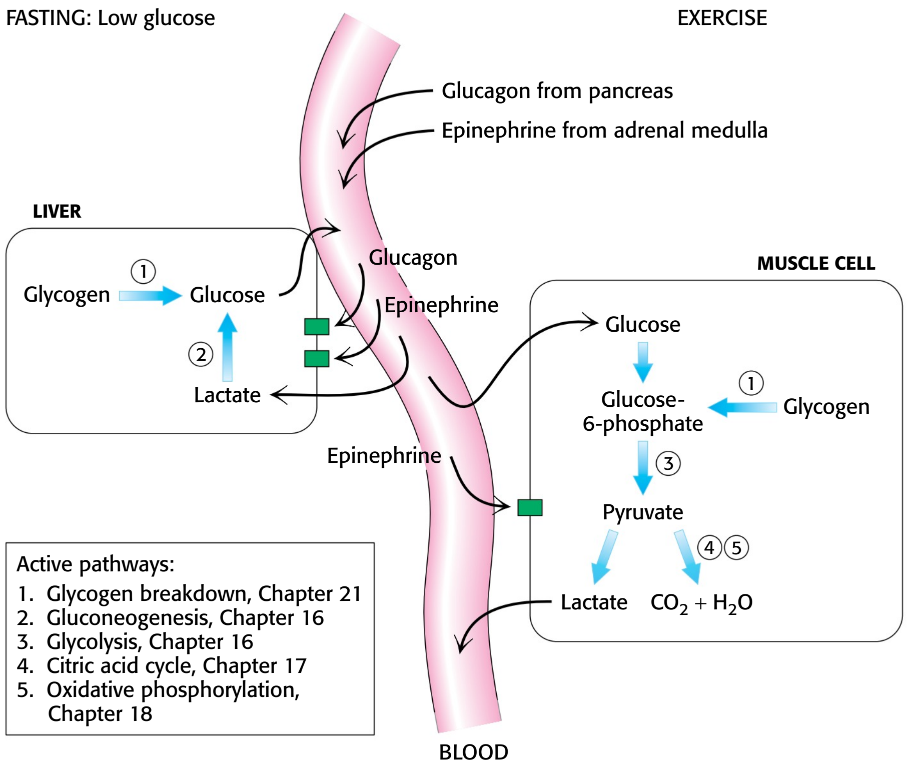

FASTING: Low glucose
EXERCISE
Glucagon from pancreas
Epinephrine from adrenal medulla
LIVER
Glycogen
①
Glucose
②
Lactate
Glucagon
Epinephrine
MUSCLE CELL
Glucose
Glucose-6-phosphate
①
Glycogen
③
Pyruvate
④⑤
Lactate
CO₂ + H₂O
Epinephrine
Active pathways:
1. Glycogen breakdown, Chapter 21
2. Gluconeogenesis, Chapter 16
3. Glycolysis, Chapter 16
4. Citric acid cycle, Chapter 17
5. Oxidative phosphorylation, Chapter 18
BLOOD

FIGURE 21.15 PATHWAY INTEGRATION: Hormonal control of glycogen breakdown. Glucagon stimulates liver-glycogen breakdown when blood glucose is low. Epinephrine enhances glycogen breakdown in muscle and the liver to provide fuel for muscle contraction.

2. The GTP-bound subunit of G_s activates the transmembrane protein adenylate cyclase. This enzyme catalyzes the formation of the second messenger cyclic AMP (cAMP) from ATP.

3. The elevated cytoplasmic level of cAMP activates *protein kinase A* (Section 10.3). The binding of cAMP to inhibitory regulatory subunits

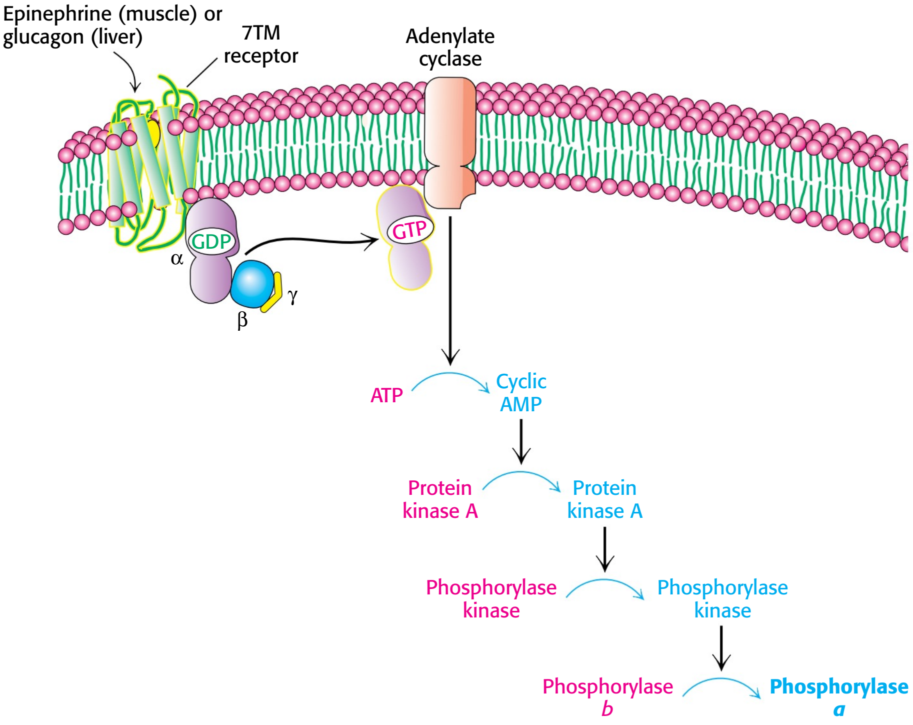

Epinephrine (muscle) or glucagon (liver)
7TM receptor
Adenylate cyclase
GDP
α
β
γ
GTP
ATP
Cyclic AMP
Protein kinase A
Protein kinase A
Phosphorylase kinase
Phosphorylase kinase
Phosphorylase b
Phosphorylase a

FIGURE 21.16 Regulatory cascade for glycogen breakdown. Glycogen degradation is stimulated by hormone binding to 7TM receptors. Hormone binding initiates a G-protein-dependent signal-transduction pathway that results in the phosphorylation and activation of glycogen phosphorylase.

triggers their dissociation from the catalytic subunits. The free catalytic subunits are now active.

4. Protein kinase A phosphorylates phosphorylase kinase, first on the β subunit and then on the α subunit, which activates glycogen phosphorylase.

The cyclic AMP cascade highly amplifies the effects of hormones. The binding of a small number of hormone molecules to cell-surface receptors leads to the release of a very large number of sugar units. Indeed, much of the stored glycogen would be mobilized within seconds were it not for a counterregulatory system.

The signal-transduction processes in the liver are more complex than those in muscle. Epinephrine can also elicit glycogen degradation in the liver. However, in addition to binding to the β-adrenergic receptor, it binds to the 7TM α-adrenergic receptor, which then initiates the *phosphoinositide cascade* (Section 14.2) that induces the release of Ca^{2+} from endoplasmic reticulum stores. Recall that the δ subunit of phosphorylase kinase is the Ca^{2+} sensor calmodulin. The binding of Ca^{2+} to calmodulin leads to a partial activation of phosphorylase kinase. Stimulation by both glucagon and epinephrine leads to maximal mobilization of liver glycogen.

Glycogen breakdown must be rapidly turned off when necessary

It is crucial that the high-gain system of glycogen breakdown be terminated quickly to prevent the wasteful depletion of glycogen after energy needs have been met. When glucose needs have been satisfied, phosphorylase kinase and glycogen phosphorylase are dephosphorylated and inactivated. Simultaneously, glycogen synthesis is activated.

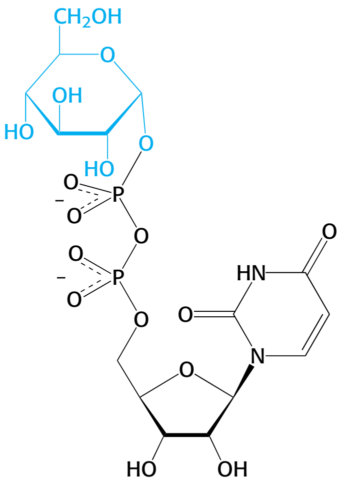
The signal-transduction pathway leading to the activation of glycogen phosphorylase is shut down automatically when secretion of the initiating hormone ceases. The inherent GTPase activity of the G protein converts the bound GTP into inactive GDP, and phosphodiesterases always present in the cell convert cyclic AMP into AMP. Protein phosphatase 1 (PP1) removes the phosphoryl groups from phosphorylase kinase, thereby inactivating the enzyme. Finally, PP1 also removes the phosphoryl group from glycogen phosphorylase, converting the enzyme into the usually inactive *b* form.

The regulation of glycogen phosphorylase became more sophisticated as the enzyme evolved

Analyses of the primary structures of glycogen phosphorylase from human beings, rats, *Dictyostelium* (slime mold), yeast, potatoes, and *E. coli* have enabled inferences to be made about the evolution of this important enzyme. The 16 residues that come into contact with glucose at the active site are identical in nearly all the enzymes. There is more variation but still substantial conservation of the 15 residues at the pyridoxal phosphate-binding site. Likewise, the glycogen-binding site is well conserved in all the enzymes. The high degree of similarity among these three sites shows that the catalytic mechanism has been maintained throughout evolution.

Differences arise, however, when we compare the regulatory sites. The simplest type of regulation would be feedback inhibition by glucose 6-phosphate. Indeed, the glucose 6-phosphate regulatory site is highly conserved among most of the phosphorylases. The crucial amino acid residues that participate in regulation by phosphorylation and nucleotide binding are well conserved only in the mammalian enzymes. Thus, this level of regulation was a later evolutionary acquisition.

21.4 Glycogen Is Synthesized and Degraded by Different Pathways


Uridine diphosphate glucose (UDP-glucose)

As with glycolysis and gluconeogenesis, biosynthetic and degradative pathways rarely operate by precisely the same reactions in the forward and reverse directions. Glycogen metabolism provided the first known example of this important principle. *Separate pathways afford much greater flexibility, both in energetics and in control.*

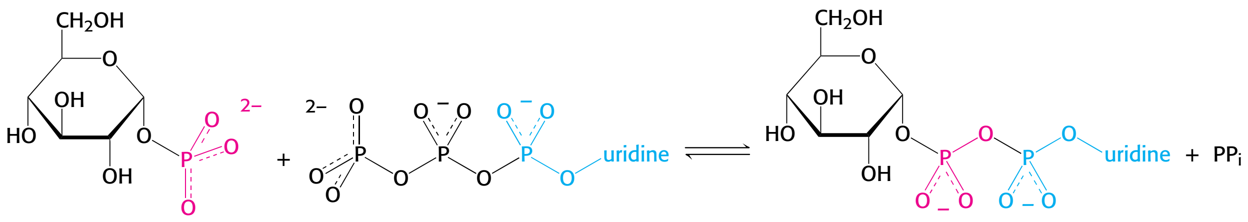
Glycogen is synthesized by a pathway that utilizes *uridine diphosphate glucose* (UDP-glucose) as the activated glucose donor.

$$\text{Synthesis: Glycogen}_n + \text{UDP-glucose} \longrightarrow \text{glycogen}_{n+1} + \text{UDP}$$
$$\text{Degradation: Glycogen}_{n+1} + \text{P}_i \longrightarrow \text{glycogen}_n + \text{glucose 1-phosphate}$$

UDP-glucose is an activated form of glucose

UDP-glucose, the glucose donor in the biosynthesis of glycogen, is an *activated form of glucose,* just as ATP and acetyl CoA are activated forms of orthophosphate and acetate, respectively. The C-1 carbon atom of the glucosyl unit of UDP-glucose is activated because its hydroxyl group is esterified to the diphosphate moiety of UDP.

UDP-glucose is synthesized from glucose 1-phosphate and uridine triphosphate (UTP) in a reaction catalyzed by *UDP-glucose pyrophosphorylase.* This reaction liberates the outer two phosphoryl residues of UTP as pyrophosphate.


Glucose 1-phosphate
UTP
UDP-glucose

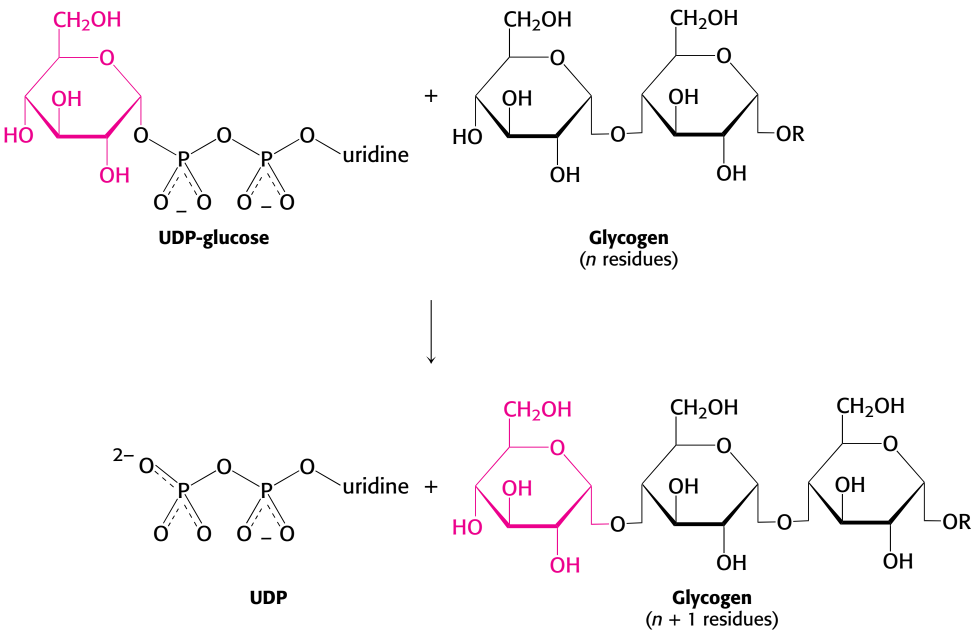
This reaction is readily reversible. However, pyrophosphate is rapidly hydrolyzed in vivo to orthophosphate by an inorganic pyrophosphatase. The essentially irreversible hydrolysis of pyrophosphate drives the synthesis of UDP-glucose.

$$\begin{array}{r}
\text{Glucose 1-phosphate} + \text{UTP} \rightleftharpoons \text{UDP-glucose} + \text{PP}_i \\
\text{PP}_i + \text{H}_2\text{O} \longrightarrow 2\,\text{P}_i \\
\hline
\text{Glucose 1-phosphate} + \text{UTP} + \text{H}_2\text{O} \longrightarrow \text{UDP-glucose} + 2\,\text{P}_i
\end{array}$$

The synthesis of UDP-glucose exemplifies another recurring theme in biochemistry: *Many biosynthetic reactions are driven by the hydrolysis of pyrophosphate.*

Glycogen synthase catalyzes the transfer of glucose from UDP-glucose to a growing chain

New glucosyl units are added to the nonreducing terminal residues of glycogen. The activated glucosyl unit of UDP-glucose is transferred to the hydroxyl group at C-4 of a terminal residue to form an α-1,4-glycosidic linkage. UDP is displaced by the terminal hydroxyl group of the growing glycogen molecule. This reaction is catalyzed by *glycogen synthase, the key regulatory enzyme in glycogen synthesis.* Humans have two isozymes of glycogen synthase: one is specific to the liver while the other is expressed in muscle and other tissues.


CH₂OH
O
OH
HO
OH
O
P
O
P
O
uridine
O – O
O – O
UDP-glucose
+
CH₂OH
CH₂OH
HO
OH
OH
OR
Glycogen
(n residues)
2–
UDP
Glycogen
(n + 1 residues)

Glycogen synthase, a member of the large glycosyltransferase family (Section 11.3), can add glucosyl residues only to a polysaccharide chain already containing at least four residues. Thus, glycogen synthesis requires a *primer*. This priming function is carried out by *glycogenin*, a Mn^{2+}-requiring glycosyltransferase composed of two identical 37-kDa subunits. Each subunit of glycogenin catalyzes the formation of α-1,4-glucose polymers, 10 to 20 glucosyl units in length, which are covalently attached to the phenolic hydroxyl group of a specific tyrosine residue in each glycogenin subunit. UDP-glucose is the donor in this autoglycosylation. At this point, glycogen synthase takes over to extend the glycogen molecule. Thus, every glycogen molecule has a glycogenin molecule at its core (Figure 21.1).

Despite no detectable sequence similarity, structural studies have revealed that glycogen synthase is homologous to glycogen phosphorylase. The binding site for UDP-glucose in glycogen synthase corresponds in position to the pyridoxal phosphate in glycogen phosphorylase.

A branching enzyme forms α-1,6 linkages

Glycogen synthase catalyzes only the synthesis of α-1,4 linkages. Another enzyme is required to form the α-1,6 linkages that make glycogen a branched polymer. Branching takes place after a number of glucosyl residues are joined in α-1,4 linkages by glycogen synthase (Figure 21.17). A branch is created by the breaking of an α-1,4 link and the formation of an α-1,6 link. A block of residues, typically 7 in number, is transferred to a more interior site. The *branching enzyme* that catalyzes this reaction requires that the block of 7 or so residues must include the nonreducing

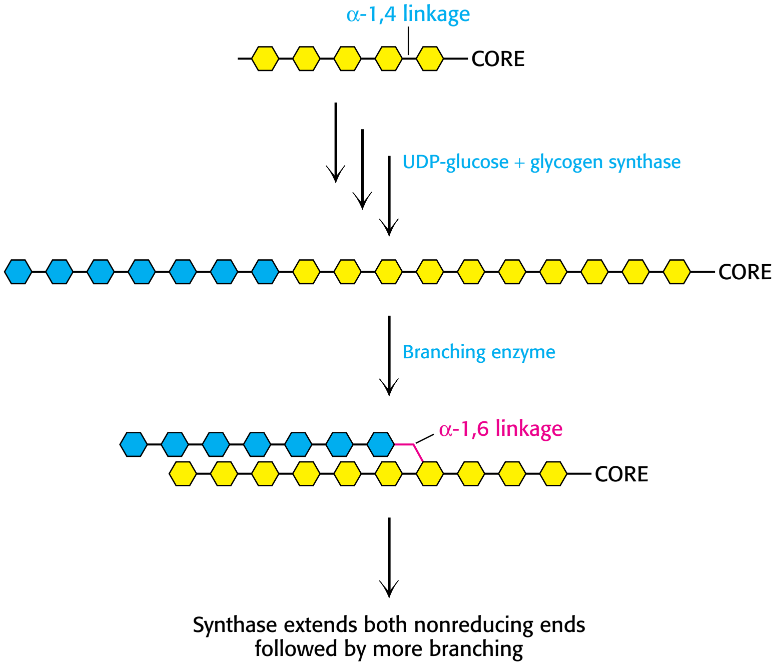

α-1,4 linkage
CORE
UDP-glucose + glycogen synthase
CORE
Branching enzyme
α-1,6 linkage
CORE
Synthase extends both nonreducing ends
followed by more branching

FIGURE 21.17 Branching reaction. The branching enzyme removes an oligosaccharide of approximately seven residues from the nonreducing end and creates an internal α-1,6 linkage.

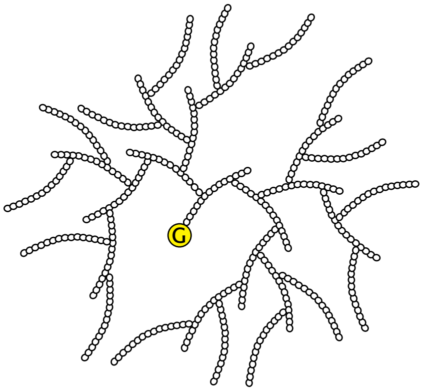


FIGURE 21.18 Cross section of a glycogen molecule. The component labeled G is glycogenin.

terminus, and must come from a chain at least 11 residues long. In addition, the new branch point must be at least 4 residues away from a preexisting one.

Branching is important because it increases the solubility of glycogen. Furthermore, branching creates a large number of terminal residues, the sites of action of glycogen phosphorylase and synthase (Figure 21.18). Thus, *branching increases the rate of glycogen synthesis and degradation.*

Glycogen synthase is the key regulatory enzyme in glycogen synthesis

Glycogen synthase, like phosphorylase, exists in two forms: an active nonphosphorylated *a* form and a usually inactive phosphorylated *b* form. Again, like the phosphorylase, the interconversion of the two forms is regulated by covalent modification. However, the key means of regulating glycogen synthase is by allosteric regulation of the phosphorylated form of the enzyme, glycogen synthase *b*. Glucose 6-phosphate is a powerful activator of the enzyme, stabilizing the R state of the enzyme relative to the T state.

The covalent modification of glycogen synthase appears to play more of a fine-tuning role. The synthase is phosphorylated at multiple sites by several protein kinases—notably, *glycogen synthase kinase* (GSK), which is under the control of insulin (p. 635), and *protein kinase A*. The function of the multiple phosphorylation sites is still under investigation. Note that *phosphorylation has opposite effects on the enzymatic activities of glycogen synthase and glycogen phosphorylase.*

Glycogen is an efficient storage form of glucose

What is the cost of converting glucose 6-phosphate into glycogen and back into glucose 6-phosphate? The pertinent reactions have already been described, except for reaction 5, which is the regeneration of UTP. ATP phosphorylates UDP in a reaction catalyzed by *nucleoside diphosphokinase.*

$$\text{Glucose 6-phosphate} \longrightarrow \text{glucose 1-phosphate} \tag{1}$$

$$\text{Glucose 1-phosphate} + \text{UTP} \longrightarrow \text{UDP-glucose} + \text{PP}_i \tag{2}$$

$$\text{PP}_i + \text{H}_2\text{O} \longrightarrow 2\ \text{P}_i \tag{3}$$

$$\text{UDP-glucose} + \text{glycogen}_n \longrightarrow \text{glycogen}_{n+1} + \text{UDP} \tag{4}$$

$$\text{UDP} + \text{ATP} \longrightarrow \text{UTP} + \text{ADP} \tag{5}$$

$$\text{Sum: Glucose 6-phosphate} + \text{ATP} + \text{glycogen}_n + \text{H}_2\text{O} \longrightarrow \text{glycogen}_{n+1} + \text{ADP} + 2\ \text{P}_i$$

Thus, 1 molecule of ATP is hydrolyzed to incorporate glucose 6-phosphate into glycogen. The energy yield from the breakdown of glycogen is highly efficient. About 90% of the residues are phosphorolytically cleaved to glucose 1-phosphate, which is converted at no cost into glucose 6-phosphate. The other residues are branch residues, which are hydrolytically cleaved. One molecule of ATP is then used to phosphorylate each of these glucose molecules to glucose 6-phosphate. The complete oxidation of glucose 6-phosphate yields about 31 molecules of ATP, and storage consumes slightly more than 1 molecule of ATP per molecule of glucose 6-phosphate; so *the overall efficiency of storage is nearly 97%.*

21.5 Glycogen Breakdown and Synthesis Are Reciprocally Regulated

An important control mechanism prevents glycogen from being synthesized at the same time as it is being broken down. *The same glucagon- and epinephrine-triggered cAMP cascades that initiate glycogen breakdown in the*

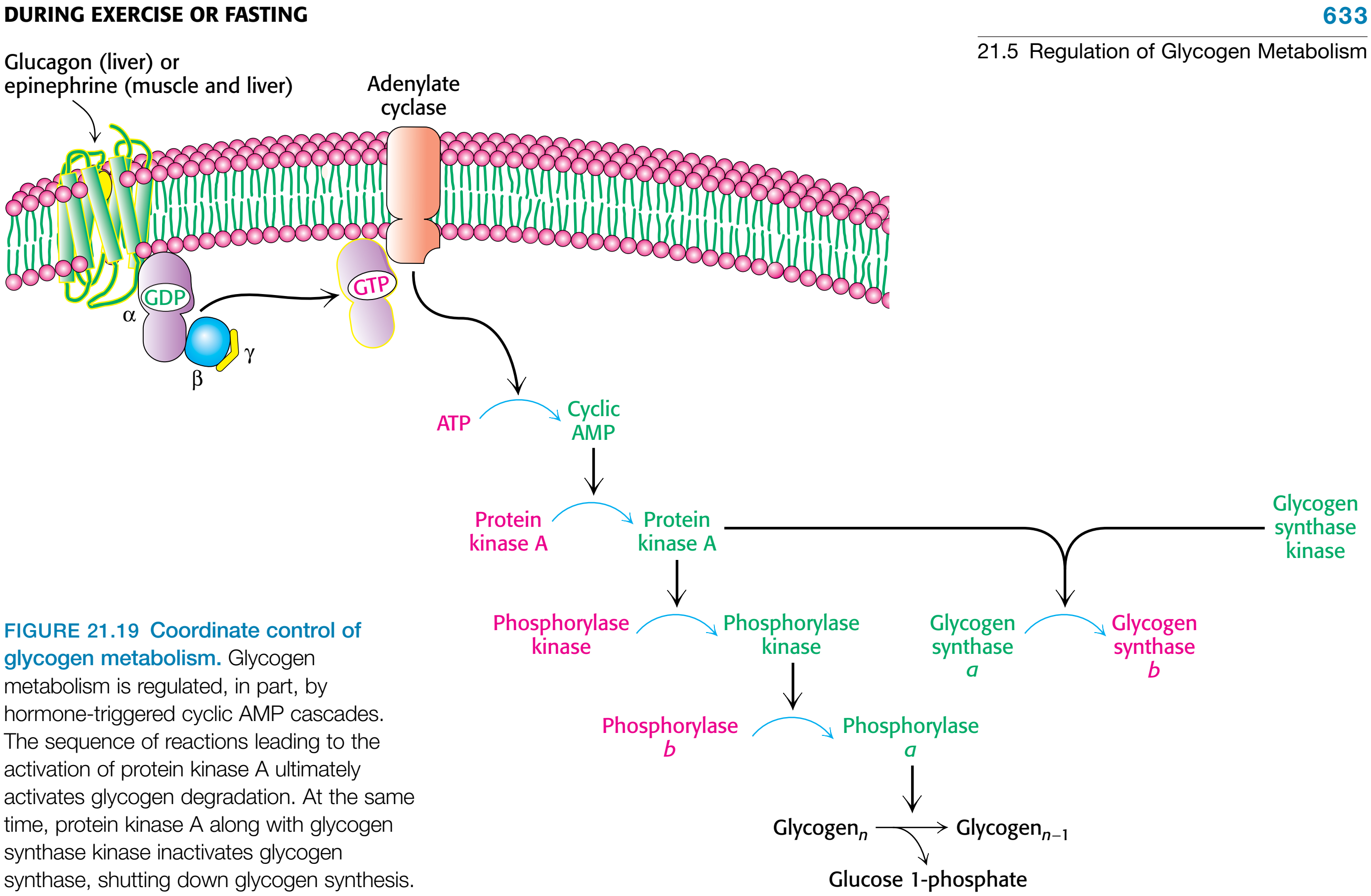

DURING EXERCISE OR FASTING
Glucagon (liver) or epinephrine (muscle and liver)
Adenylate cyclase
GDP
GTP
α
β
γ
ATP
Cyclic AMP
Protein kinase A
Protein kinase A
Glycogen synthase kinase
Phosphorylase kinase
Phosphorylase kinase
Glycogen synthase a
Glycogen synthase b
Phosphorylase b
Phosphorylase a
Glycogen$_n$
Glycogen$_{n-1}$
Glucose 1-phosphate

FIGURE 21.19 Coordinate control of glycogen metabolism. Glycogen metabolism is regulated, in part, by hormone-triggered cyclic AMP cascades. The sequence of reactions leading to the activation of protein kinase A ultimately activates glycogen degradation. At the same time, protein kinase A along with glycogen synthase kinase inactivates glycogen synthase, shutting down glycogen synthesis.

liver and muscle, respectively, also shut off glycogen synthesis. Glucagon and epinephrine control both glycogen breakdown and glycogen synthesis through protein kinase A (Figure 21.19). Recall that protein kinase A adds a phosphoryl group to phosphorylase kinase, activating that enzyme and initiating glycogen breakdown. Glycogen synthase kinase and protein kinase A add phosphoryl groups to glycogen synthase, but this phosphorylation leads to a *decrease* in enzymatic activity. In this way, glycogen breakdown and synthesis are reciprocally regulated. How is the enzymatic activity reversed so that glycogen breakdown halts and glycogen synthesis begins?

Protein phosphatase 1 reverses the regulatory effects of kinases on glycogen metabolism

After a bout of exercise, muscle must shift from a glycogen-degrading mode to one of glycogen replenishment. A first step in this metabolic task is to shut down the phosphorylated proteins that stimulate glycogen breakdown. This task is accomplished by *protein phosphatases* that catalyze the hydrolysis of phosphorylated serine and threonine residues in proteins. *Protein phosphatase 1 plays key roles in regulating glycogen metabolism* (Figure 21.20). PP1 inactivates phosphorylase *a* and phosphorylase kinase by dephosphorylating them. PP1 decreases the rate of glycogen breakdown; it reverses the effects of the phosphorylation cascade. Moreover, *PP1 also removes phosphoryl groups from glycogen synthase* b *to convert it into the more active glycogen synthase* a *form.* Here, PP1 also accelerates glycogen synthesis. PP1 is yet another molecular device for coordinating carbohydrate storage. Interestingly, recent research suggests that glycogen phosphorylase is also regulated by acetylation (Chapter 10.3). Acetylation not only inhbits the enzyme, but also enhances dephosphorylation by promoting the interaction with PP1. Acetylation is stimulated

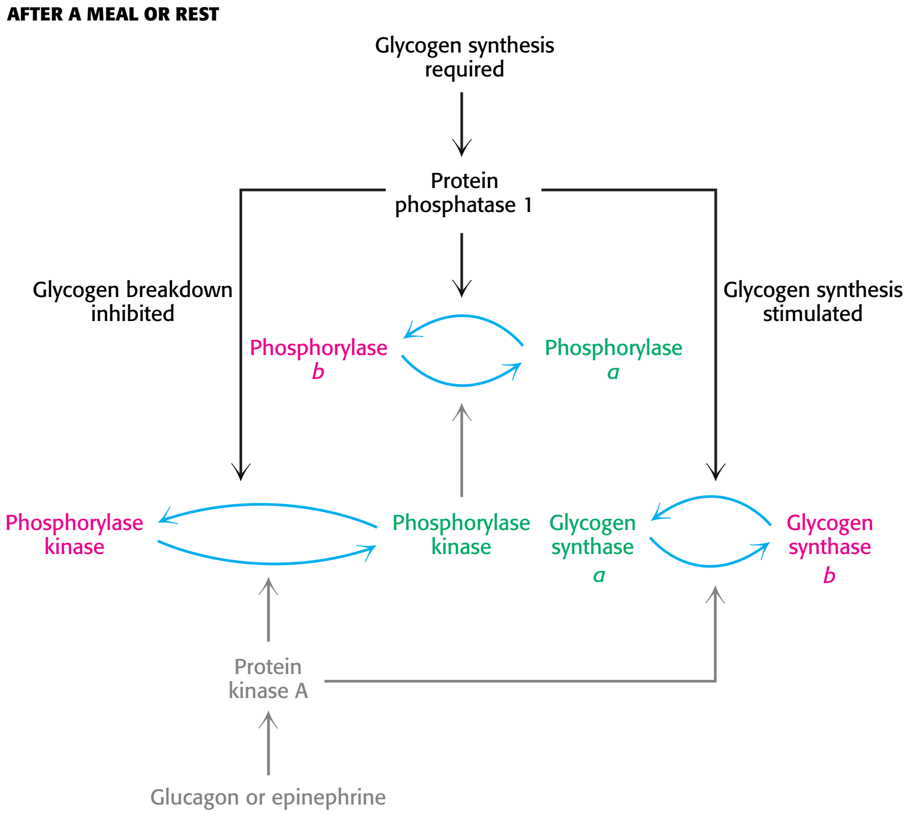
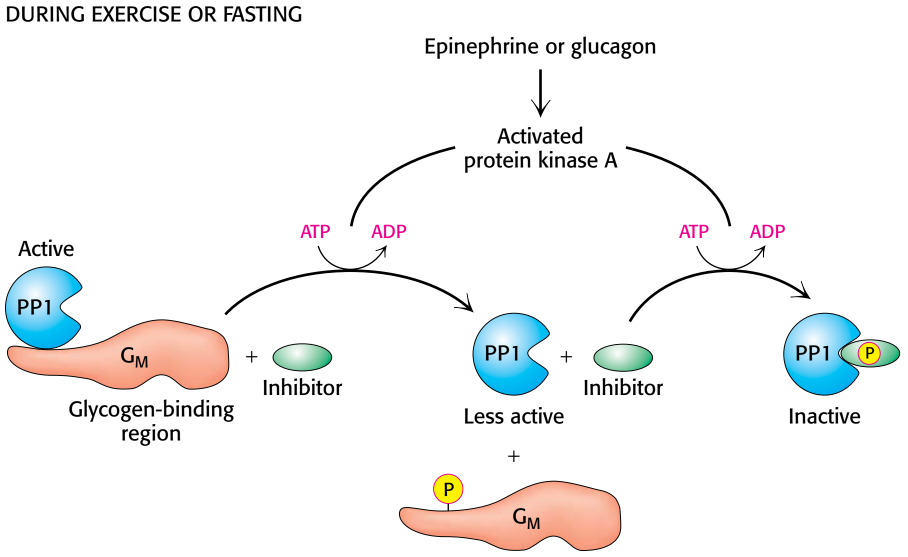

AFTER A MEAL OR REST
Glycogen synthesis required
Protein phosphatase 1
Glycogen breakdown inhibited
Glycogen synthesis stimulated
Phosphorylase b
Phosphorylase a
Phosphorylase kinase
Phosphorylase kinase
Glycogen synthase a
Glycogen synthase b
Protein kinase A
Glucagon or epinephrine

FIGURE 21.20 Regulation of glycogen synthesis by protein phosphatase 1. Protein phosphatase 1 stimulates glycogen synthesis while inhibiting glycogen breakdown. Active enzymes are shown in green and inactive enzymes in red.

by glucose and insulin, but inhibited by glucagon. It is fascinating that glycogen phosphorylase, one of the first allosteric enzymes identified and one of the most studied enzymes, is still revealing secrets.

The catalytic subunit of PP1 is a 37-kDa single-domain protein. This subunit is usually bound to one of a family of regulatory subunits with masses of approximately 120 kDa. In skeletal muscle and heart, the most prevalent regulatory subunit is called G_M, whereas in the liver, the most prevalent subunit is G_L. These regulatory subunits have modular structures with domains that participate in interactions with glycogen, with the catalytic subunit, and with target enzymes. Thus, *these regulatory subunits act as scaffolds, bringing together the phosphatase and its substrates on the glycogen particle.*

FIGURE 21.21 Regulation of protein phosphatase 1 (PP1) in muscle takes place in two steps. Phosphorylation of G_M by protein kinase A dissociates the catalytic subunit from its substrates in the glycogen particle. Phosphorylation of the inhibitor subunit by protein kinase A inactivates the catalytic unit of PP1.


DURING EXERCISE OR FASTING
Epinephrine or glucagon
Activated protein kinase A
ATP
ADP
ATP
ADP
Active
PP1
G_M
Glycogen-binding region
+
Inhibitor
PP1
Less active
+
Inhibitor
+
P
G_M
PP1
P
Inactive

What prevents the phosphatase activity of PP1 from always inhibiting glycogen degradation? When glycogen degradation is called for, epinephrine or glucagon initiate the cAMP cascade that activates protein kinase A (Figure 21.21). Protein kinase A reduces the activity of PP1 by two mechanisms. First, in muscle, G_M is phosphorylated in the domain responsible for binding the catalytic subunit. The catalytic subunit is released from glycogen and from its substrates and dephosphoryation is greatly reduced. Second, almost all tissues contain small proteins that, when phosphorylated, bind to the catalytic

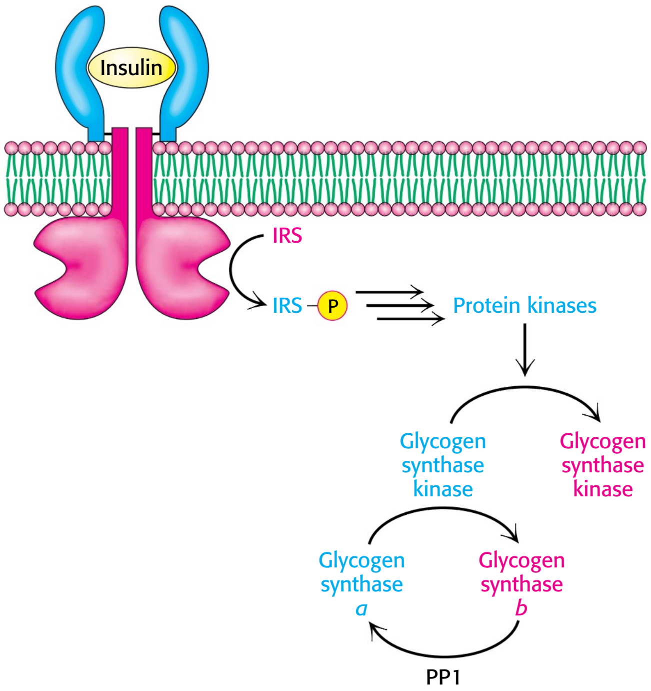


FIGURE 21.22 Insulin inactivates glycogen synthase kinase. Insulin triggers a cascade that leads to the phosphorylation and inactivation of glycogen synthase kinase and prevents the phosphorylation of glycogen synthase. Protein phosphatase 1 (PP1) removes the phosphates from glycogen synthase, thereby activating the enzyme and allowing glycogen synthesis. IRS, insulin-receptor substrate.

subunit of PP1 and inhibit it. Thus, when glycogen degradation is switched on by cAMP, the accompanying phosphorylation of these inhibitors keeps phosphorylase in its active *a* form and glycogen synthase in its inactive *b* form.

Insulin stimulates glycogen synthesis by inactivating glycogen synthase kinase

After exercise, people often consume carbohydrate-rich foods to restock their glycogen stores. How is glycogen synthesis stimulated? When blood-glucose levels are high, *insulin stimulates the synthesis of glycogen by inactivating glycogen synthase kinase,* one of the enzymes that maintains glycogen synthase in its phosphorylated, inactive state (Figure 21.22). The first step in the action of insulin is its binding to its receptor, a tyrosine kinase receptor in the plasma membrane (Section 14.2). The binding of insulin stimulates the tyrosine kinase activity of the receptor so that it phosphorylates insulin-receptor substrates (IRSs). These phosphorylated proteins trigger signal-transduction pathways that eventually lead to the activation of protein kinases that phosphorylate and inactivate glycogen synthase kinase. The inactive kinase can no longer maintain glycogen synthase in its phosphorylated, inactive state. Protein phosphatase 1 dephosphorylates glycogen synthase, activating it, and restoring glycogen reserves. Recall that insulin increases the amount of glucose in the cell by increasing the number of glucose transporters in the membrane (Section 16.2). The net effect of insulin is thus the replenishment of glycogen stores.

Glycogen metabolism in the liver regulates the blood-glucose level

After a meal rich in carbohydrates, blood-glucose levels rise, and glycogen synthesis is stepped up in the liver. Although insulin is the primary signal for glycogen synthesis, another is the concentration of glucose in the blood, which normally ranges from about 4.4–6.7 mM. The liver senses the concentration of glucose in the blood and takes up or releases glucose

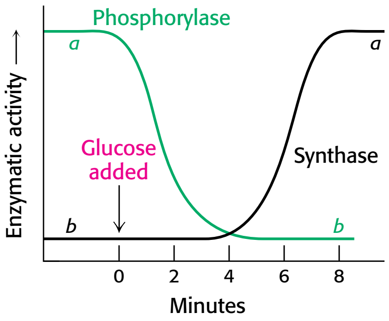


FIGURE 21.23 Blood glucose regulates liver-glycogen metabolism. The infusion of glucose into the bloodstream leads to the inactivation of phosphorylase, followed by the activation of glycogen synthase, in the liver. [Data from W. Stalmans, H. De Wulf, L. Hue, and H.-G. Hers. *Eur. J. Biochem.* 41:117–134, 1974.]

accordingly. The amount of liver phosphorylase *a* decreases rapidly when glucose is infused (Figure 21.23). After a lag period, the amount of glycogen synthase *a* increases, which results in glycogen synthesis. *Phosphorylase* a *is the glucose sensor in liver cells,* facilitating the switch from degradation to synthesis. How does phosphorylase perform its sensor function?

Phosphorylase *a* and PP1 are localized to the glycogen particle by interactions with the G_L subunit of PP1. The binding of glucose to phosphorylase *a* shifts its allosteric equilibrium from the active R form to the inactive T form. This conformational change *renders the phosphoryl group on serine 14 a substrate for protein phosphatase 1.* PP1 binds tightly to phosphorylase *a* only when the phosphorylase is in the R state but is inactive when bound. When glucose induces the transition to the T form, PP1 and the phosphorylase dissociate from each other and the glycogen particle, and PP1 becomes active, converting the phosphorylase *a* form to the *b* form. Recall that the R ↔ T transition of muscle phosphorylase *a* is unaffected by glucose and is thus unaffected by the rise in blood-glucose levels (Section 21.2).

How does glucose binding to glycogen phosphorylase stimulate glycogen synthesis? As mentioned above, the conversion of *a* into *b* is accompanied by the *release of PP1, which is then free to activate glycogen synthase* (Figure 21.24). The removal of the phosphoryl group of inactive glycogen synthase *b* converts it into the active *a* form. What accounts for the lag between termination of glycogen degradation and the beginnings of glycogen synthesis (Figure 21.23)? There are about 10 phosphorylase *a* molecules per molecule of phosphatase. Consequently, *the activity of glycogen synthase begins to increase only after most of phosphorylase* a *is converted into* b. The lag between the decrease in glycogen degradation and the increase in glycogen synthesis prevents the two pathways from operating simultaneously. This remarkable glucose-sensing system depends on three key elements: (1) communication between the allosteric site for glucose and the serine phosphate, (2) the use of PP1 to inactivate phosphorylase and activate glycogen synthase, and (3) the binding of the phosphatase to phosphorylase *a* to prevent the premature activation of glycogen synthase.

Efforts are underway to develop drugs that disrupt the interaction of liver phosphorylase with the G_L subunit as a treatment for type 2 diabetes (Section 27.2). One experimental drug binds at a unique site on the enzyme that, synergistically with glucose, stabilizes the T state of

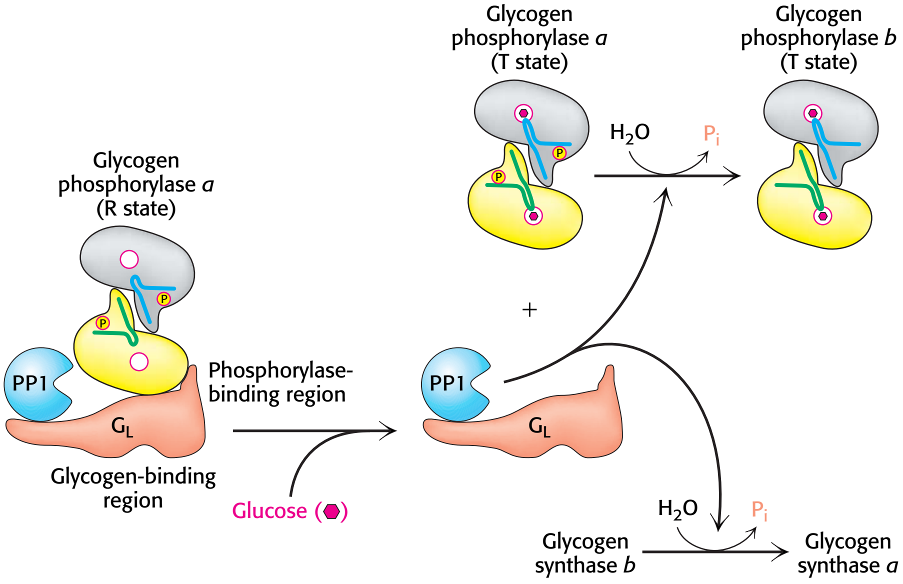

Glycogen phosphorylase *a* (R state)
P
P
PP1
G_L
Phosphorylase-binding region
Glycogen-binding region
Glucose (●)
Glycogen phosphorylase *a* (T state)
P
P
H_2O
P_i
Glycogen phosphorylase *b* (T state)
+
PP1
G_L
H_2O
P_i
Glycogen synthase *b*
Glycogen synthase *a*

FIGURE 21.24 Glucose regulation of liver-glycogen metabolism. Glucose binds to and inhibits glycogen phosphorylase *a* in the liver, facilitating the formation of the T state of phosphorylase *a*. The T state of phosphorylase *a* does not bind protein phosphate 1 (PP1), leading to the dissociation and activation of PP1 from glycogen phosphorylase *a*. The free PP1 dephosphorylates glycogen phosphorylase *a* and glycogen synthase *b*, leading to the inactivation of glycogen breakdown and the activation of glycogen synthesis.

phosphorylase and thereby enhances the transition to glycogen synthesis, as described earlier. Type 2 diabetes is characterized by excess blood glucose. Hence, disrupting the association of phosphorylase with the G_L would render it a substrate for PP1. Glycogen degradation would cease, and glucose release into the blood would be inhibited.

A biochemical understanding of glycogen-storage diseases is possible

Edgar von Gierke described the first glycogen-storage disease in 1929. A patient with this disease has a huge abdomen caused by a massive enlargement of the liver. There is a pronounced hypoglycemia between meals. Furthermore, the blood-glucose level does not rise on administration of epinephrine and glucagon. An infant with this glycogen-storage disease may have convulsions because of the low blood-glucose level.

The enzymatic defect in von Gierke disease was elucidated in 1952 by Carl and Gerty Cori. They found that *glucose 6-phosphatase is missing from the liver of a patient with this disease.* This finding was the first demonstration of an inherited deficiency of a liver enzyme. The glycogen in the liver is normal in structure but is present in abnormally large amounts. The absence of glucose 6-phosphatase in the liver causes hypoglycemia because glucose cannot be formed from glucose 6-phosphate. This phosphorylated sugar does not leave the liver, because it cannot cross the plasma membrane. The presence of excess glucose 6-phosphate triggers an increase in glycolysis in the liver, leading to a high level of lactate and pyruvate in the blood. Patients who have von Gierke disease also have an increased dependence on fat metabolism. This disease can also be produced by a mutation in the gene that encodes the *glucose 6-phosphate transporter.* Recall that glucose 6-phosphate must be transported into the lumen of the endoplasmic reticulum to be hydrolyzed by phosphatase (Section 16.3). Mutations in the other three essential proteins of this system can likewise lead to von Gierke disease.

Seven other glycogen-storage diseases have been characterized (Table 21.2). In Pompe disease (type II), lysosomes become engorged with

TABLE 21.2 Glycogen-storage diseases

Type	Defective enzyme	Organ affected	Glycogen in the affected organ	Clinical features
I Von Gierke	Glucose 6-phosphatase or transport system	Liver and kidney	Increased amount; normal structure.	Massive enlargement of the liver; failure to thrive; severe hypoglycemia, ketosis, hyperuricemia, hyperlipemia.
II Pompe	α-1,4-Glucosidase (lysosomal)	All organs	Massive increase in amount; normal structure.	Cardiorespiratory failure causes death, usually before age 2.
III Cori	α-1,6-Glucosidase (debranching enzyme)	Muscle and liver	Increased amount; short outer branches.	Like type I, but milder course.
IV Andersen	Branching enzyme (α-1,4 $\longrightarrow$ α-1,6)	Liver and spleen	Normal amount; very long outer branches.	Progressive cirrhosis of the liver; liver failure causes death, usually before age 2.
V McArdle	Phosphorylase	Muscle	Moderately increased amount; normal structure.	Limited ability to perform strenuous exercise because of painful muscle cramps; otherwise patient is normal and well developed.
VI Hers	Phosphorylase	Liver	Increased amount.	Like type I, but milder course.
VII	Phosphofructokinase	Muscle	Increased amount; normal structure.	Like type V.
VIII	Phosphorylase kinase	Liver	Increased amount; normal structure.	Mild liver enlargement; mild hypoglycemia.

Note: Types I through VII are inherited as autosomal recessives. Type VIII is sex linked.

glycogen because they lack α-1,4-glucosidase, a hydrolytic enzyme confined to these organelles (Figure 21.25). Carl and Gerty Cori also elucidated the biochemical defect in another glycogen-storage disease (type III), which cannot be distinguished from von Gierke disease (type I) by physical examination alone. In type III disease, the structure of liver and muscle glycogen is abnormal and the amount is markedly increased. Most striking, the outer branches of the glycogen are very short. *Patients having this type lack the debranching enzyme (α-1,6-glucosidase),* and so only the outermost branches of glycogen can be effectively utilized. Thus, only a small fraction of this abnormal glycogen is functionally active as an accessible store of glucose.

A defect in glycogen metabolism confined to muscle is found in McArdle disease (type V). *Muscle phosphorylase activity is absent,* and a patient's capacity to perform strenuous exercise is limited because of painful muscle cramps. The patient is otherwise normal and well developed. Thus, effective utilization of muscle glycogen is not essential for life.

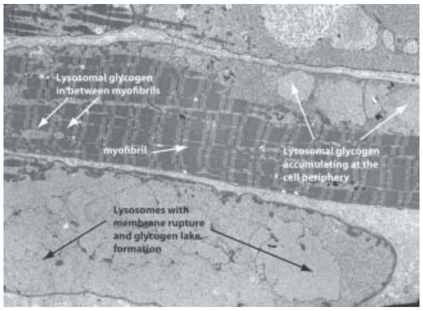

Lysosomal glycogen in between myofibrils
myofibril
Lysosomal glycogen accumulating at the cell periphery
Lysosomes with membrane rupture and glycogen lake formation

FIGURE 21.25 Electron micrograph of muscle cell from a patient with Pompe disease. Glycogen-engorged lysosomes are seen throughout the cell, including the myofibrils. As the disease progresses, lysosomes may rupture, releasing large amounts of glycogen into the cytoplasm. Such accumulations of cytoplasmic glycogen are called glycogen lakes. [Reproduced with permission of the author, from Thurberg, B. L. et. al. Characterization of pre- and post-treatment pathology after enyme replacement for Pompe disease. *Lab. Invest.* 86 (12):1208–1220, 2006.]

SUMMARY

Glycogen, a readily mobilized fuel store, is a branched polymer of glucose residues. Most of the glucose units in glycogen are linked by α-1,4-glycosidic bonds. At about every twelfth residue, a branch is created by an α-1,6-glycosidic bond. Glycogen is present in large amounts in muscle cells and in liver cells, where it is stored in the cytoplasm in the form of hydrated granules.

21.1 Glycogen Breakdown Requires the Interplay of Several Enzymes

Most of the glycogen molecule is degraded to glucose 1-phosphate by the action of glycogen phosphorylase, the key enzyme in glycogen breakdown. The glycosidic linkage between C-1 of a terminal residue and C-4 of the adjacent one is split by orthophosphate to give glucose 1-phosphate, which can be reversibly converted into glucose 6-phosphate. Branch points are degraded by the concerted action of an oligosaccharide transferase and an α-1,6-glucosidase.

21.2 Phosphorylase Is Regulated by Allosteric Interactions and Reversible Phosphorylation

Phosphorylase *b,* which is usually inactive, is converted into active phosphorylase *a* by the phosphorylation of a single serine residue in each subunit. This reaction is catalyzed by phosphorylase kinase. The *a* form in the liver is inhibited by glucose. In liver, phosphorylase is activated to liberate glucose for export to other organs, such as skeletal muscle and the brain. The *b* form in muscle can be activated by the binding of AMP, an effect counteracted by ATP and glucose 6-phosphate. In contrast to liver, muscle phosphorylase is activated to generate glucose for use inside the cell as a fuel for contractile activity.

21.3 Epinephrine and Glucagon Signal the Need for Glycogen Breakdown

Epinephrine and glucagon stimulate glycogen breakdown through specific 7TM receptors. Muscle is the primary target of epinephrine, whereas the liver is responsive to glucagon. Both signal molecules initiate a kinase cascade that leads to the activation of phosphorylase kinase, which in turn converts glycogen phosphorylase *b* to the phosphorylated *a* form.

21.4 Glycogen Is Synthesized and Degraded by Different Pathways

The pathway for glycogen synthesis differs from that for glycogen breakdown. UDP-glucose, the activated intermediate in glycogen synthesis, is formed from glucose 1-phosphate and UTP. Glycogen synthase catalyzes the transfer of glucose from UDP-glucose to the C-4 hydroxyl group of a terminal residue in the growing glycogen molecule. Synthesis is primed by glycogenin, an autoglycosylating protein that contains a covalently attached oligosaccharide unit on a specific tyrosine residue. A branching enzyme converts some of the α-1,4 linkages into α-1,6 linkages to increase the number of ends so that glycogen can be made and degraded more rapidly.

21.5 Glycogen Breakdown and Synthesis Are Reciprocally Regulated

Glycogen synthesis and degradation are coordinated by several amplifying reaction cascades. Epinephrine and glucagon stimulate glycogen breakdown and inhibit its synthesis by increasing the cytoplasmic level of cyclic AMP, which activates protein kinase A. Protein kinase A activates glycogen breakdown by attaching a phosphate to phosphorylase kinase and inhibits glycogen synthesis by phosphorylating glycogen synthase. Glycogen synthase kinase also inhibits synthesis by phosphorylating the synthase.

The glycogen-mobilizing actions of protein kinase A are reversed by protein phosphatase 1, which is regulated by several hormones. Epinephrine inhibits this phosphatase by blocking its attachment to glycogen molecules and by turning on an inhibitor. Insulin, in contrast, triggers a cascade that phosphorylates and inactivates glycogen synthase kinase. Hence, glycogen synthesis is decreased by epinephrine and increased by insulin. Glycogen synthase and phosphorylase are also regulated by noncovalent allosteric interactions. In fact, phosphorylase is a key part of the glucose-sensing system of liver cells. Glycogen metabolism exemplifies the power and precision of reversible phosphorylation in regulating biological processes.

KEY TERMS

glycogen phosphorylase (p. 619)
phosphorolysis (p. 619)
pyridoxal phosphate (PLP) (p. 620)
phosphorylase kinase (p. 626)
calmodulin (p. 626)
epinephrine (adrenaline) (p. 627)
glucagon (p. 627)
protein kinase A (PKA) (p. 628)
uridine diphosphate glucose (UDP-glucose) (p. 630)
glycogen synthase (p. 630)
glycogenin (p. 631)
glycogen synthase kinase (p. 632)
protein phosphatase 1 (PP1) (p. 633)
insulin (p. 635)

PROBLEMS

1. *Step-by-step degradation.* What are the three steps in glycogen degradation and what enzymes are required?

2. *Match 'em.* Match each term with its description.

(a) Glycogen phosphorylase ____
(b) Phosphorolysis ____
(c) Transferase ____
(d) α-1,6-Glucosidase ____
(e) Phosphoglucomutase ____
(f) Phosphorylase kinase ____
(g) Protein kinase A ____
(h) Calmodulin ____
(i) Epinephrine ____
(j) Glucagon ____

1. Calcium-binding subunit of phosphorylase kinase
2. Activates glycogen phosphorylase
3. Removal of a glucose residue by the addition of phosphate
4. Stimulates glycogen breakdown in muscle
5. Liberates a free glucose residue
6. Shifts the location of several glucose residues
7. Stimulates glycogen breakdown in the liver
8. Catalyzes phosphorolytic cleavage
9. Prepares glucose 1-phosphate for glycolysis
10. Phosphorylates phosphorylase kinase

3. *Choice is good*. Glycogen is not as reduced as fatty acids are and consequently not as energy rich. Why do animals store any energy as glycogen? Why not convert all excess fuel into fatty acids?

4. *If a little is good, a lot is better*. α-Amylose is an unbranched glucose polymer. Why would this polymer not be as effective a storage form of glucose as glycogen?

5. *Telltale products*. A sample of glycogen from a patient with liver disease is incubated with orthophosphate, phosphorylase, the transferase, and the debranching enzyme (α-1,6-glucosidase). The ratio of glucose 1-phosphate to glucose formed in this mixture is 100. What is the most likely enzymatic deficiency in this patient?

6. *Dare to be different*. Compare the allosteric regulation of phosphorylase in the liver and in muscle, and explain the significance of the difference.

7. *A thumb on the balance*. The reaction catalyzed by phosphorylase is readily reversible in vitro. At pH 6.8, the equilibrium ratio of orthophosphate to glucose 1-phosphate is 3.6. The value of $\Delta G^{\circ\prime}$ for this reaction is small because a glycosidic bond is replaced by a phosphoryl ester bond that has a nearly equal transfer potential. However, phosphorolysis proceeds far in the direction of glycogen breakdown in vivo. Suggest one means by which the reaction can be made irreversible in vivo.

8. *Excessive storage*. Suggest an explanation for the fact that the amount of glycogen in type I glycogen-storage disease (von Gierke disease) is increased.

9. *Recouping an essential phosphoryl*. The phosphoryl group on phosphoglucomutase is slowly lost by hydrolysis. Propose a mechanism that utilizes a known catalytic intermediate for restoring this essential phosphoryl group. How might this phosphoryl donor be formed?

10. *Not all absences are equal*. Hers disease results from an absence of liver glycogen phosphorylase and may result in serious illness. In McArdle disease, muscle glycogen phosphorylase is absent. Although exercise is difficult for patients suffering from McArdle disease, the disease is rarely life threatening. Account for the different manifestations of the absence of glycogen phosphorylase in the two tissues. What does the existence of these two different diseases indicate about the genetic nature of the phosphorylase?

11. *Hydrophobia*. Why is water excluded from the active site of phosphorylase? Predict the effect of a mutation that allows water molecules to enter.

12. *Removing all traces*. In human liver extracts, the catalytic activity of glycogenin was detectable only after treatment with α-amylase, an enzyme the hydrolyzes α-1,4-bonds in starch and glycogen. Why was α-amylase necessary to reveal the glycogenin activity?

13. *Two in one*. A single polypeptide chain houses the transferase and debranching enzyme. Cite a potential advantage of this arrangement.

14. *How did they do that?* A strain of mice has been developed that lack the enzyme phosphorylase kinase. Yet, after strenuous exercise, the glycogen stores of a mouse of this strain are depleted. Explain how this depletion is possible.

15. *An appropriate inhibitor*. What is the rationale for the inhibition of muscle glycogen phosphorylase by glucose 6-phosphate when glucose 1-phosphate is the product of the phosphorylase reaction?

16. *Passing along the information*. Outline the signal-transduction cascade for glycogen degradation in muscle.

17. *Slammin' on the breaks*. There must be a way to shut down glycogen breakdown quickly to prevent the wasteful depletion of glycogen after energy needs have been met. What mechanisms are employed to turn off glycogen breakdown?

18. *Diametrically opposed*. Phosphorylation has opposite effects on glycogen synthesis and breakdown. What is the advantage of its having opposing effects?

19. *Feeling depleted*. Glycogen depletion resulting from intense, extensive exercise can lead to exhaustion and the inability to continue exercising. Some people also experience dizziness, an inability to concentrate, and a loss of muscle control. Account for these symptoms.

20. *Everyone had a job to do*. What accounts for the fact that liver phosphorylase is a glucose sensor, whereas muscle phosphorylase is not?

21. *Yin and Yang*. Match the terms on the left with the descriptions on the right.

(a) UDP-glucose ____
(b) UDP-glucose pyrophosphorylase____
(c) Glycogen synthase ____
(d) Glycogenin____
(e) Branching enzyme ____
(f) Glucose 6-phosphate ____
(g) Glycogen synthase kinase ____

1. Glucose 1-phosphate is one of its substrates.
2. Potent activator of glycogen synthase *b*.
3. Glucose sensor in the liver.
4. Activated substrate for glycogen synthesis.
5. Synthesizes α-1,4 linkages between glucose molecules.
6. Leads to the inactivation of glycogen synthase kinase.
7. Synthesizes α-1,6 linkages between glucose molecules.

(h) Protein phosphatase 1 ____
(i) Insulin ____
(j) Glycogen phosphorylase *a* ____

8. Catalyzes the formation of glycogen synthase *b*.
9. Catalyzes the formation of glycogen synthase *a*.
10. Provides the primer for glycogen synthesis.

22. *Team effort*. What enzymes are required for the synthesis of a glycogen particle starting from glucose 6-phosphate?

23. *Force it forward*. The following reaction accounts for the synthesis of UDP-glucose. This reaction is readily reversible. How is it made irreversible in vivo?

$$\text{Glucose 1-phosphate} + \text{UTP} \rightleftharpoons \text{UDP-glucose} + \text{PP}_i$$

24. *If you insist*. Why does activation of the phosphorylated *b* form of glycogen synthase by high concentrations of glucose 6-phosphate make good biochemical sense?

25. *An ATP saved is an ATP earned*. The complete oxidation of glucose 6-phosphate derived from free glucose yields 30 molecules ATP, whereas the complete oxidation of glucose 6-phosphate derived from glycogen yields 31 molecules of ATP. Account for this difference.

26. *Dual roles*. Phosphoglucomutase is crucial for glycogen breakdown as well as for glycogen synthesis. Explain the role of this enzyme in each of the two processes.

27. *Working at cross-purposes*. Write a balanced equation showing the effect of simultaneous activation of glycogen phosphorylase and glycogen synthase. Include the reactions catalyzed by phosphoglucomutase and UDP-glucose pyrophosphorylase.

28. *Achieving immortality*. Glycogen synthase requires a primer. A primer was formerly thought to be provided when the existing glycogen granules are divided between the daughter cells produced by cell division. In other words, parts of the original glycogen molecule were simply passed from generation to generation. Would this strategy have been successful in passing glycogen stores from generation to generation? How are new glycogen molecules now known to be synthesized?

29. *Synthesis signal*. How does insulin stimulate glycogen synthesis?

Mechanism Problem

30. *Family resemblance*. Enzymes of the α-amylase family (problem 12) catalyze a reaction by forming a covalent intermediate to a conserved aspartate residue. Propose mechanisms for the two enzymes catalyzing steps in glycogen debranching on the basis of their potential membership in the α-amylase family.

Chapter Integration Problems

31. *Double activation*. What pathway in addition to the cAMP-induced signal transduction pathway is used in the liver to maximize glycogen breakdown?

32. *Carbohydrate conversion*. Write a balanced equation for the formation of glycogen from galactose.

33. *Working together*. What enzymes are required for the liver to release glucose into the blood when an organism is asleep and fasting?

34. *A shattering experience*. Crystals of phosphorylase *a* grown in the presence of glucose shatter when a substrate such as glucose 1-phosphate is added. Why?

35. *I know I've seen that face before*. UDP-glucose is the activated form of glucose used in glycogen synthesis. However, we have previously met other similar activated forms of carbohydrate in our consideration of metabolism. Where else have we seen UDP-carbohydrate?

36. *Same symptoms, different cause*. Suggest another mutation in glucose metabolism that causes symptoms similar to those of von Gierke disease.

Data Interpretation Problems

37. *An authentic replica*. Experiments were performed in which serine (S) 14 of glycogen phosphorylase was replaced by glutamate (E). The V_{max} of the mutant enzyme was then compared with the wild type phosphorylase in both the *a* and the *b* form.

	V_{max} μmoles of glucose 1-PO_4 released min^{-1} mg^{-1}
Wild type phosphorylase *b*	25 ± 0.4
Wild type phosphorylase *a*	100 ± 5
S to E mutant	60 ± 3

(a) Explain the results obtained with the mutant.

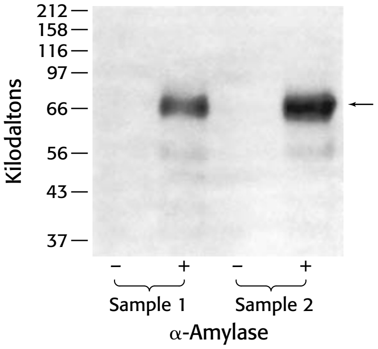
(b) Predict the effect of substituting aspartic acid for the serine.

38. *Glycogen isolation 1*. The liver is a major storage site for glycogen. Purified from two samples of human liver, glycogen was either treated or not treated with α-amylase and subsequently analyzed by SDS-PAGE and western blotting with the use of antibodies to glycogenin. The results are presented in the adjoining illustration on the next page.

(a) Why are no proteins visible in the lanes without amylase treatment?

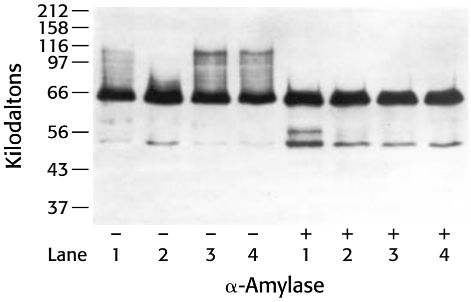
(b) What is the effect of treating the samples with α-amylase? Explain the results.

(c) List other proteins that you might expect to be associated with glycogen. Why are other proteins not visible?



[Courtesy of Dr. Peter J. Roach, Indiana University School of Medicine.]

39. *Glycogen isolation 2.* The gene for glycogenin was transfected into a cell line that normally stores only small amounts of glycogen. The cells were then manipulated according to the following protocol, and glycogen was isolated and analyzed by SDS-PAGE and western blotting by using an antibody to glycogenin with and without α-amylase treatment. The results are presented in the adjoining illustration.



[Courtesy of Dr. Peter J. Roach, Indiana University School of Medicine.]

The protocol: Cells cultured in growth medium and 25 mM glucose (lane 1) were switched to medium containing no glucose for 24 hours (lane 2). Glucose-starved cells were refed with medium containing 25 mM glucose for 1 hour (lane 3) or 3 hours (lane 4). Samples (12 μg of protein) were either treated or not treated with α-amylase, as indicated, before being loaded on the gel.

(a) Why did the western analysis produce a "smear"—that is, the high-molecular-weight staining in lane 1(−)?

(b) What is the significance of the decrease in high-molecular-weight staining in lane 2(−)?

(c) What is the significance of the difference between lanes 2(−) and 3(−)?

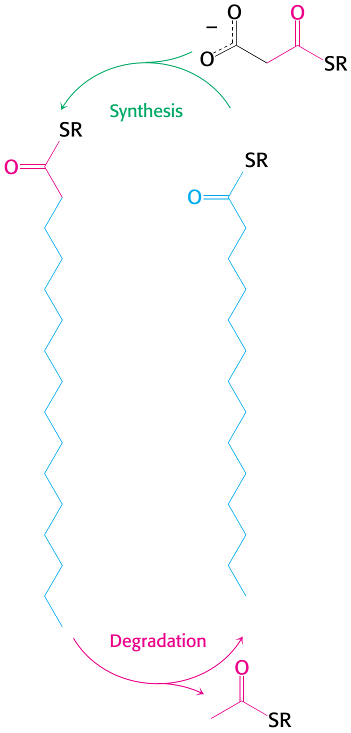
(d) Suggest a plausible reason why there is essentially no difference between lanes 3(−) and 4(−).

(e) Why are the bands at 66 kDa the same in the lanes treated with amylase, despite the fact that the cells were treated differently?

CHAPTER

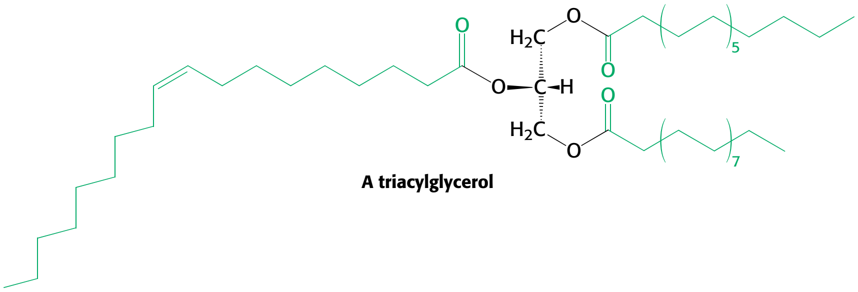
22

Fatty Acid Metabolism



Fats provide efficient means for storing energy for later use. (Right) The processes of fatty acid synthesis (preparation for energy storage) and fatty acid degradation (preparation for energy use) are, in many ways, the reverse of each other. (Above) Studies of mice are revealing the interplay between these pathways and the biochemical bases of appetite and weight control. [Photograph Oak Ridge National Laboratory/U.S. Department of Energy/Science Photo Library.]

We turn now from the metabolism of carbohydrates to that of fatty acids. A fatty acid contains a long hydrocarbon chain and a terminal carboxylate group. Fatty acids have four major physiological roles. First, *fatty acids are fuel molecules.* They are stored as *triacylglycerols* (also called *neutral fats* or *triglycerides*), which are uncharged esters of fatty acids with glycerol. Triacylglycerols are stored in adipose tissue, composed of cells called adipocytes (Figure 22.1). Fatty acids mobilized from triacylglycerols are oxidized to meet the energy needs of a cell or organism. During rest or moderate exercise, such as walking, fatty acids are our primary source of energy. Second, *fatty acids are building blocks of phospholipids and glycolipids.* These amphipathic molecules are important components of biological membranes, as discussed in Chapter 12. Third, many proteins are modified by the *covalent attachment of fatty acids, which targets the proteins to membrane locations* (Section 12.4). Fourth, *fatty acid derivatives serve as hormones and intracellular messengers.* In this chapter, we focus on the degradation and synthesis of fatty acids.



A triacylglycerol

OUTLINE

22.1 Triacylglycerols Are Highly Concentrated Energy Stores

22.2 The Use of Fatty Acids as Fuel Requires Three Stages of Processing

22.3 Unsaturated and Odd-Chain Fatty Acids Require Additional Steps for Degradation

22.4 Fatty Acids Are Synthesized by Fatty Acid Synthase

22.5 The Elongation and Unsaturation of Fatty Acids Are Accomplished by Accessory Enzyme Systems

22.6 Acetyl CoA Carboxylase Plays a Key Role in Controlling Fatty Acid Metabolism

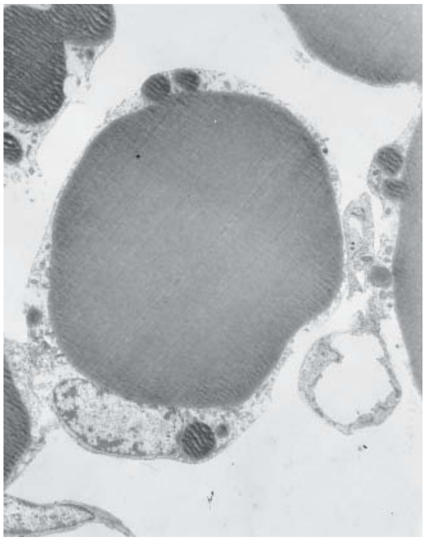


FIGURE 22.1 **Electron micrograph of an adipocyte.** A small band of cytoplasm surrounds the large deposit of triacylglycerols. [Biophoto Associates/Photo Researchers.]

Fatty acid degradation and synthesis mirror each other in their chemical reactions

Fatty acid degradation and synthesis consist of four steps that are the reverse of each other in their basic chemistry. Degradation is an oxidative process that converts a fatty acid into a set of activated acetyl units (acetyl CoA) that can be processed by the citric acid cycle (Figure 22.2). An activated fatty acid is oxidized to introduce a double bond; the double bond is hydrated to introduce a hydroxyl group; the alcohol is oxidized to a ketone; and, finally, the fatty acid is cleaved by coenzyme A to yield acetyl CoA and a fatty acid chain two carbons shorter. If the fatty acid has an even number of carbon atoms and is saturated, the process is simply repeated until the fatty acid is completely converted into acetyl CoA units.

Fatty acid synthesis is essentially the reverse of this process. The process starts with the individual units to be assembled—in this case, with an activated acyl group (most simply, an acetyl unit) and a malonyl unit (Figure 22.2). The malonyl unit condenses with the acetyl unit to form a four-carbon fragment. To produce the required hydrocarbon chain, the carbonyl group is reduced to a methylene group in three steps: a reduction,

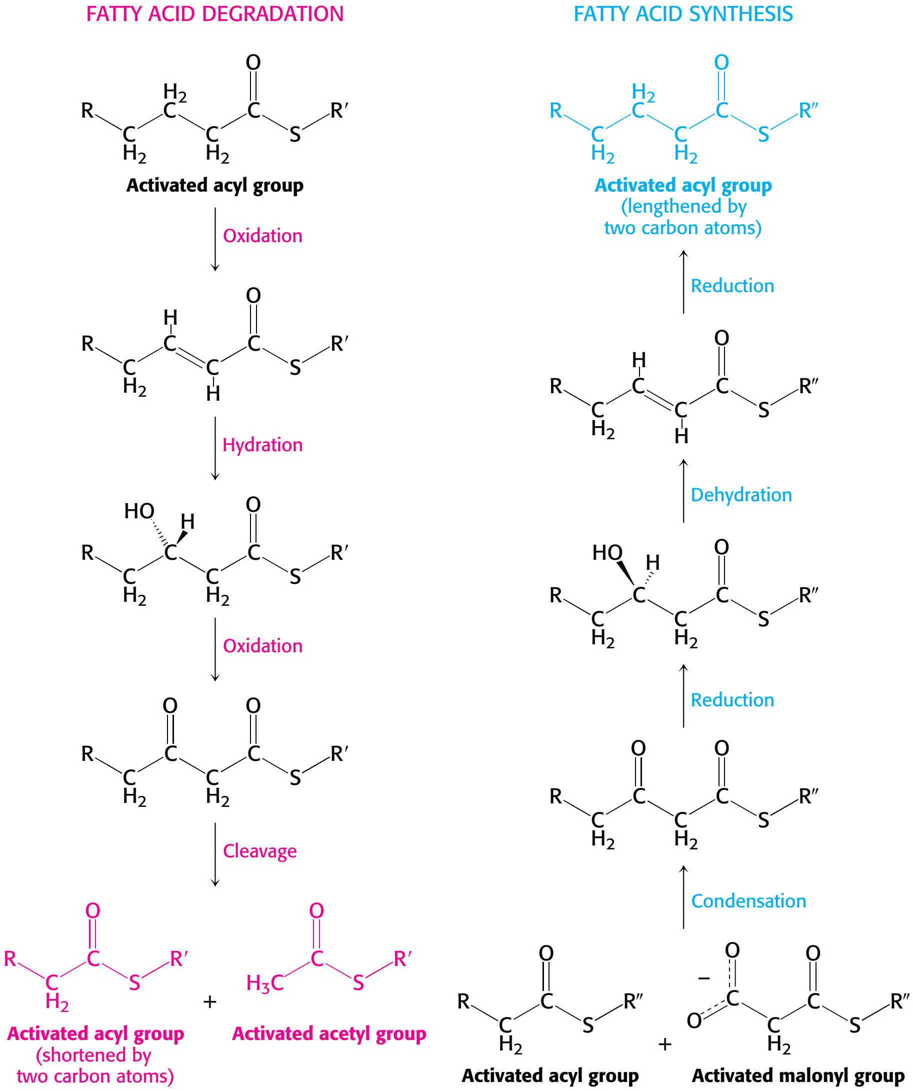

FATTY ACID DEGRADATION
FATTY ACID SYNTHESIS
Activated acyl group
Oxidation
Hydration
Oxidation
Cleavage
Activated acyl group (shortened by two carbon atoms)
Activated acetyl group
Activated acyl group (lengthened by two carbon atoms)
Reduction
Dehydration
Reduction
Condensation
Activated acyl group
Activated malonyl group

FIGURE 22.2 **Steps in fatty acid degradation and synthesis.** The two processes are in many ways mirror images of each other.

a dehydration, and another reduction, exactly the opposite of degradation. The product of the reduction is butyryl CoA. Another activated malonyl group condenses with the butyryl unit, and the process is repeated until a C_{16} or shorter fatty acid is synthesized.

22.1 Triacylglycerols Are Highly Concentrated Energy Stores

Triacylglycerols are highly concentrated stores of metabolic energy because they are *reduced* and *anhydrous.* The yield from the complete oxidation of fatty acids is about 38 kJ g^{-1} (9 kcal g^{-1}), in contrast with about 17 kJ g^{-1} (4 kcal g^{-1}) for carbohydrates and proteins. The basis of this large difference in caloric yield is that fatty acids are much more reduced than carbohydrates or proteins. Furthermore, triacylglycerols are nonpolar, and so they are stored in a nearly anhydrous form, whereas much more polar carbohydrates are more highly hydrated. In fact, 1 g of dry glycogen binds about 2 g of water. Consequently, *a gram of nearly anhydrous fat stores 6.75 times as much energy as a gram of hydrated glycogen,* which is likely the reason that triacylglycerols rather than glycogen were selected in evolution as the major energy reservoir. Consider a typical 70-kg man, who has fuel reserves of 420,000 kJ (100,000 kcal) in triacylglycerols, 100,000 kJ (24,000 kcal) in protein (mostly in muscle), 2500 kJ (600 kcal) in glycogen, and 170 kJ (40 kcal) in glucose. Triacylglycerols constitute about 11 kg of his total body weight. If this amount of energy were stored in glycogen, his total body weight would be 64 kg greater. The glycogen and glucose stores provide enough energy to sustain physiological function for about 24 hours, whereas the triacylglycerol stores allow survival for several weeks.

In mammals, the major site of triacylglycerol accumulation is the cytoplasm of *adipose cells (fat cells).* This fuel-rich tissue is located throughout the body, notably under the skin (subcutaneous fat) and surrounding the internal organs (visceral fat). Droplets of triacylglycerol coalesce to form a large globule, called a *lipid droplet,* which may occupy most of the cell volume (Figure 22.1). The lipid droplet is surrounded by a monolayer of phospholipids and numerous proteins required for triacylglycerol metabolism. Originally believed to be inert-lipid deposits, lipid droplets are now understood to be dynamic organelles essential for the regulation of lipid metabolism. Adipose cells are specialized for the synthesis and storage of triacylglycerols and for their mobilization into fuel molecules that are transported to other tissues by the blood. Muscle also stores triacylglycerols for its own energy needs. Indeed, triacylglycerols are evident as the "marbling" of expensive cuts of beef.

The utility of triacylglycerols as an energy source is dramatically illustrated by the abilities of migratory birds, which can fly great distances without eating after having stored energy as triacylglycerols. Examples are the American golden plover and the ruby-throated hummingbird. The golden plover flies from Alaska to the southern tip of South America; a large segment of the flight (3800 km, or 2400 miles) is over open ocean, where the birds cannot feed. The ruby-throated hummingbird flies nonstop across the Gulf of Mexico. Fatty acids provide the energy source for both these prodigious feats.

Triacylglycerols fuel the long migration flights of the American golden plover (*Pluvialis dominica*). [Source: Jim Zipp/Science Source.]

Dietary lipids are digested by pancreatic lipases

Most lipids are ingested in the form of triacylglycerols and must be degraded to fatty acids for absorption across the intestinal epithelium. Intestinal enzymes called *lipases,* secreted by the pancreas, degrade triacylglycerols to

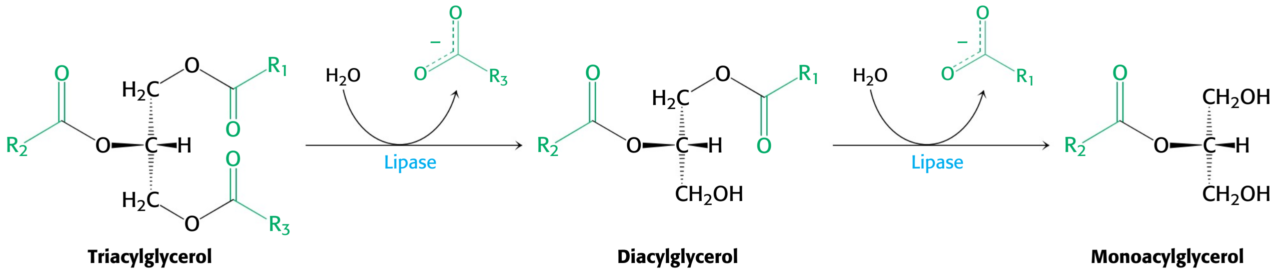


FIGURE 22.3 Action of pancreatic lipases. Lipases secreted by the pancreas convert triacylglycerols into fatty acids and monoacylglycerol for absorption into the intestine.

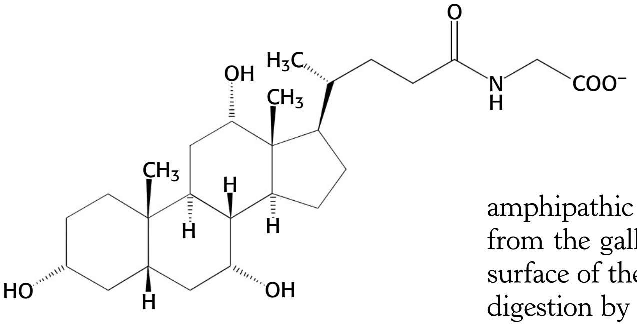


FIGURE 22.4 Glycocholate. Bile acids, such as glycocholate, facilitate lipid digestion in the intestine.

free fatty acids and monoacylglycerol (Figure 22.3). Lipids present a special problem because, unlike carbohydrates and proteins, these molecules are not soluble in water. They exit the stomach as an emulsion, particles with a triacylglycerol core surrounded by cholesterol and cholesterol esters. How are the lipids made accessible to the lipases, which are in aqueous solution? First, the particles are coated with *bile acids* (Figure 22.4), amphipathic molecules synthesized from cholesterol in the liver and secreted from the gallbladder. The ester bonds of each lipid are oriented toward the surface of the bile salt-coated particle, rendering the bond more accessible to digestion by lipases in aqueous solution. However, in this form, the particles are still not substrates for digestion. The protein colipase (another pancreatic secretory product) must bind the lipase to the particle to permit lipid degradation. The final digestion products are carried in micelles to the intestinal epithelium where they are transported across the plasma membrane (Figure 22.5). If the production of bile salts is inadequate due to liver disease, large amounts of fats (as much as 30 g day^{-1}) are excreted in the feces. This condition is referred to as *steatorrhea,* after stearic acid, a common fatty acid.

Dietary lipids are transported in chylomicrons

In the intestinal mucosal cells, the triacylglycerols are resynthesized from fatty acids and monoacylglycerols and then packaged into lipoprotein transport particles called *chylomicrons,* stable particles approximately 2000 Å (200 nm) in diameter (Figure 22.5). These particles are composed mainly of triacylglycerols, with apoliprotein B-48 (apo B-48) as the main protein component. Protein constituents of lipoprotein particles are called *apolipoproteins.* Chylomicrons also transport fat-soluble vitamins and cholesterol.

The chylomicrons are released into the lymph system and then into the blood. These particles bind to membrane-bound lipases, primarily at adipose tissue and muscle, where the triacylglycerols are once again degraded into

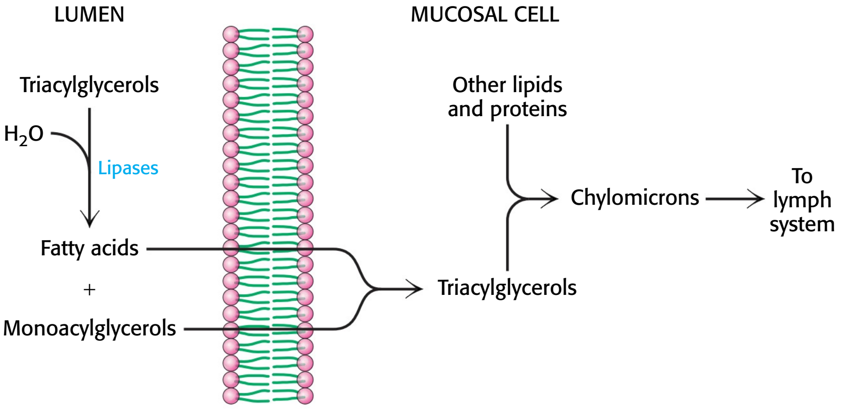
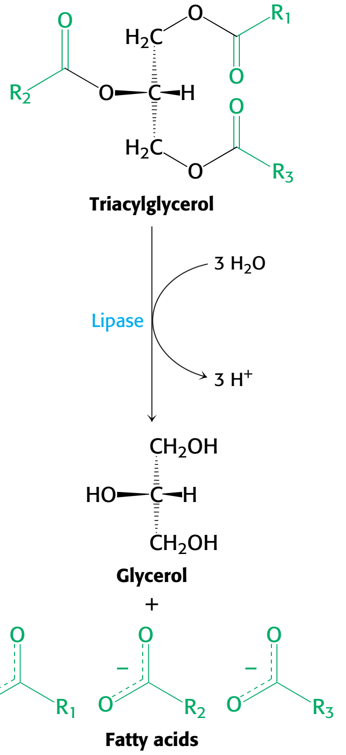


FIGURE 22.5 Chylomicron formation. Free fatty acids and monoacylglycerols are absorbed by intestinal epithelial cells. Triacylglycerols are resynthesized and packaged with other lipids and apolipoprotein B-48 to form chylomicrons, which are then released into the lymph system.

free fatty acids and monoacylglycerol for transport into the tissue. The triacylglycerols are then resynthesized inside the cell and stored. In the muscle, they can be oxidized to provide energy.

22.2 The Use of Fatty Acids as Fuel Requires Three Stages of Processing

Tissues throughout the body gain access to the lipid energy reserves stored in adipose tissue through three stages of processing. First, the lipids must be mobilized. In this process, triacylglycerols are degraded to fatty acids and glycerol, which are released from the adipose tissue and transported to the energy-requiring tissues. Second, at these tissues, the fatty acids must be activated and transported into mitochondria for degradation. Third, the fatty acids are broken down in a step-by-step fashion into acetyl CoA, which is then processed in the citric acid cycle.



Triacylglycerols are hydrolyzed by hormone-stimulated lipases

Consider someone who has just awakened from a night's sleep and begins a bout of exercise. Glycogen stores will be low, but lipids are readily available. How are these lipid stores mobilized?

Before fats can be used as fuels, the triacylglycerol storage form must be hydrolyzed to yield isolated fatty acids. This reaction is catalyzed by a hormonally controlled lipase. Under the physiological conditions facing an early-morning runner, glucagon and epinephrine will be present. In adipose tissue, these hormones trigger 7 TM receptors that activate adenylate cyclase (Section 14.1). The increased level of cyclic AMP then stimulates protein kinase A, which phosphorylates two key proteins: *perilipin,* a fat-droplet-associated protein, and hormone-sensitive lipase (Figure 22.6). The phosphorylation of perilipin has two crucial effects. First, it restructures the fat droplet so that the triacylglycerols are more accessible to the mobilization. Second, the phosphorylation of perilipin triggers the release of a coactivator for *adipose triglyceride lipase* (ATGL). Once bound to the coactivator, ATGL initiates the mobilization of triacylglycerols by releasing a fatty acid from triacylglycerol, forming diacylglycerol. Diacylglycerol is converted into a free fatty acid and

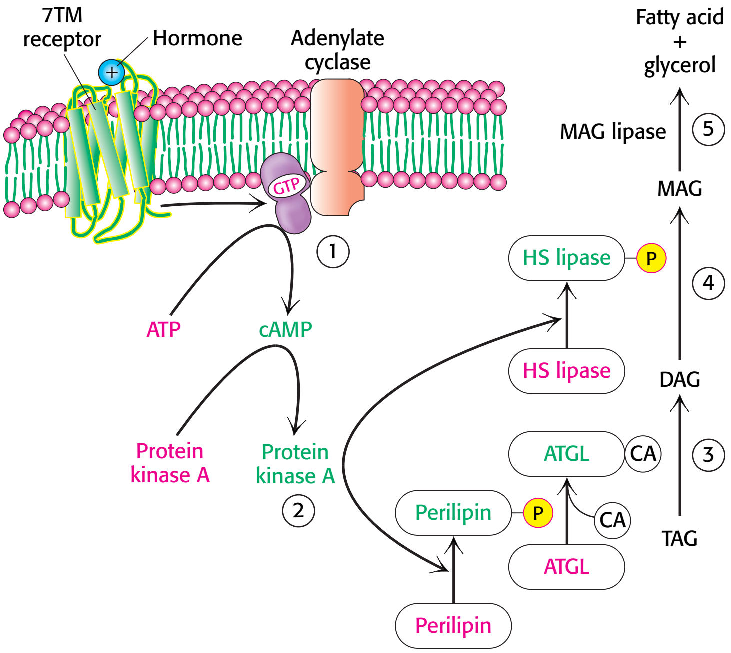


FIGURE 22.6 Mobilization of triacylglycerols. Triacylglycerols in adipose tissue are converted into free fatty acids in response to hormonal signals. 1. The hormones activate protein kinase A through the cAMP cascade. 2. Protein kinase A phosphorylates perilipin, resulting in the restructuring of the lipid droplet and release of the coactivator of ATGL, thereby activating ATGL. 3. ATGL converts triacylglycerol into diacylglycerol. 4. Hormone-sensitive lipase releases a fatty acid from diacylglycerol, generating monoacylglycerol. 5. Monoacylglycerol lipase completes the mobilization process. Abbreviations: 7TM, seven transmembrane receptor; ATGL, adipose triglyceride lipase; CA, coactivator; HS lipase, hormone-sensitive lipase; MAG lipase, monoacylglycerol lipase; DAG, diacylglycerol; TAG, triacylglycerol.

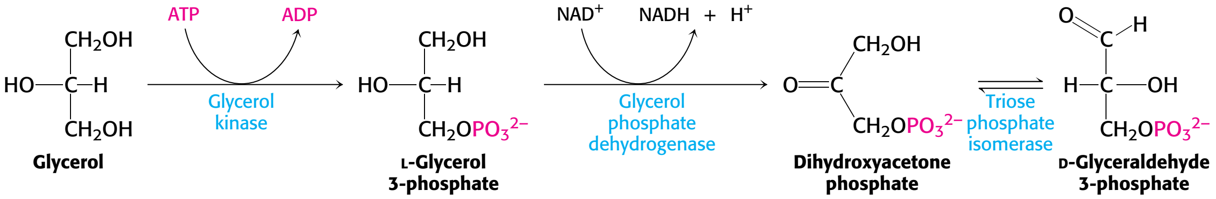
monoacylglycerol by the hormone-sensitive lipase. Finally, a monoacylglycerol lipase completes the mobilization of fatty acids with the production of a free fatty acid and glycerol. Thus, *epinephrine and glucagon induce lipolysis.* Although their role in muscle is not as firmly established, these hormones probably also regulate the use of triacylglycerol stores in that tissue.

If the coactivator required by ATGL is missing or defective, a rare condition (incidence unknown) called Chanarin-Dorfman syndrome results. Fats accumulate throughout the body because they cannot be released by ATGL. Other symptoms include dry skin (ichthyosis), enlarged liver, muscle weakness, and mild cognitive disability.

Free fatty acids and glycerol are released into the blood

Fatty acids are not soluble in aqueous solutions. In order to reach tissues that require fatty acids, the released fatty acids bind to the blood protein albumin, which delivers them to tissues in need of fuel.

Glycerol formed by lipolysis is absorbed by the liver and phosphorylated. It is then oxidized to dihydroxyacetone phosphate, which is isomerized to glyceraldehyde 3-phosphate. This molecule is an intermediate in both the glycolytic and the gluconeogenic pathways.


ATP
ADP
Glycerol kinase
NAD+
NADH + H+
Glycerol phosphate dehydrogenase
Triose phosphate isomerase
Glycerol
L-Glycerol 3-phosphate
Dihydroxyacetone phosphate
D-Glyceraldehyde 3-phosphate

Hence, glycerol can be converted into pyruvate or glucose in the liver, which contains the appropriate enzymes (Figure 22.7). The reverse process can take place by the reduction of dihydroxyacetone phosphate to glycerol 3-phosphate. Hydrolysis by a phosphatase then gives glycerol. Thus, glycerol and glycolytic intermediates are readily interconvertible.

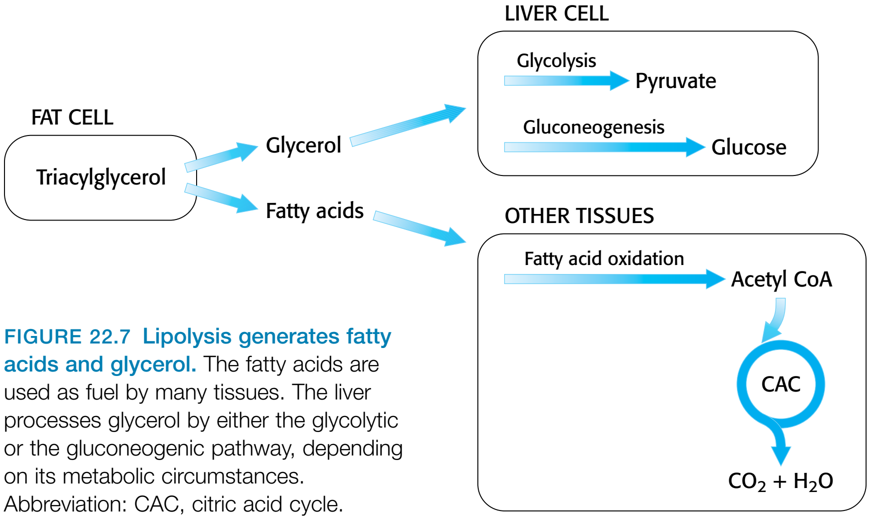
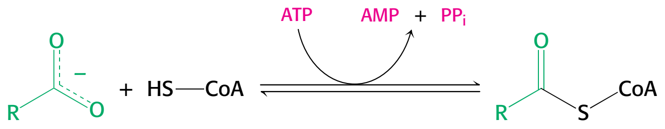

LIVER CELL
Glycolysis
Pyruvate
Gluconeogenesis
Glucose
FAT CELL
Triacylglycerol
Glycerol
Fatty acids
OTHER TISSUES
Fatty acid oxidation
Acetyl CoA
CAC
CO2 + H2O

FIGURE 22.7 Lipolysis generates fatty acids and glycerol. The fatty acids are used as fuel by many tissues. The liver processes glycerol by either the glycolytic or the gluconeogenic pathway, depending on its metabolic circumstances. Abbreviation: CAC, citric acid cycle.

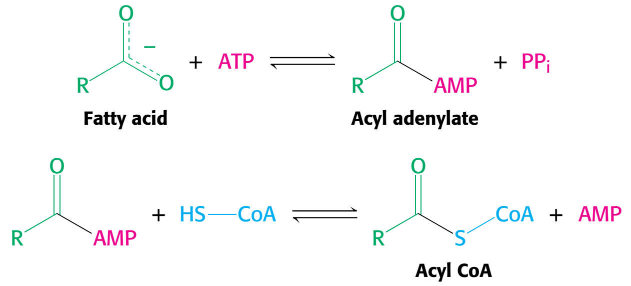
Fatty acids are linked to coenzyme A before they are oxidized

Fatty acids separate from the albumin in the blood stream and diffuse across the cell membrane with the assistance of transport proteins. In the cell, fatty acids are shuttled about in association with fatty-acid-binding-proteins.

Fatty acid oxidation occurs in the mitochondrial matrix, but in order to enter the mitochondria, the fatty acids are first activated through the formation of a thioester linkage to coenzyme A. Adenosine triphosphate (ATP) drives the formation of the thioester linkage between the carboxyl group of a fatty acid and the sulfhydryl group of coenzyme A. This activation reaction takes place on the outer mitochondrial membrane, where it is catalyzed by *acyl CoA synthetase* (also called *fatty acid thiokinase*).

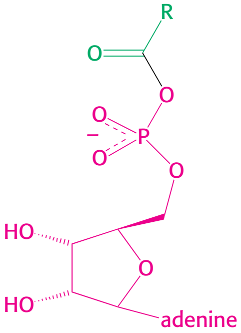


Acyl CoA synthetase accomplishes the activation of a fatty acid in two steps. First, the fatty acid reacts with ATP to form an *acyl adenylate*. In this mixed anhydride, the carboxyl group of a fatty acid is bonded to the phosphoryl group of AMP. The other two phosphoryl groups of the ATP substrate are released as pyrophosphate. In the second step, the sulfhydryl group of coenzyme A attacks the acyl adenylate, which is tightly bound to the enzyme, to form acyl CoA and AMP.



(1)

(2)



Acyl adenylate

These partial reactions are freely reversible. In fact, the equilibrium constant for the sum of these reactions is close to 1. One high-transfer-potential compound is cleaved (between PP$_i$ and AMP) and one high-transfer-potential compound is formed (the thioester acyl CoA). How is the overall reaction driven forward? The answer is that pyrophosphate is rapidly hydrolyzed by a pyrophosphatase. The complete reaction is

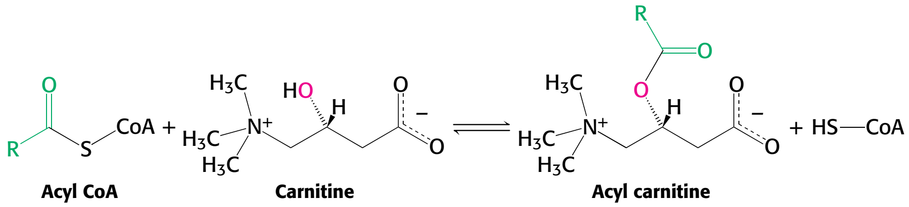
$$\mathrm{RCOO^- + CoA + ATP + H_2O \longrightarrow RCO\text{-}CoA + AMP + 2\ P_i + 2\ H^+}$$

This reaction is quite favorable because the equivalent of two molecules of ATP is hydrolyzed, whereas only one high-transfer-potential compound is formed. We see here another example of a recurring theme in biochemistry: *Many biosynthetic reactions are made irreversible by the hydrolysis of inorganic pyrophosphate.*

Another motif recurs in this activation reaction. The enzyme-bound acyl adenylate intermediate is not unique to the synthesis of acyl CoA. *Acyl adenylates are frequently formed when carboxyl groups are activated in biochemical reactions.* Amino acids are activated for protein synthesis by a similar mechanism (Section 30.2), although the enzymes that catalyze this process are not homologous to acyl CoA synthetase. Thus, *activation by adenylation recurs in part because of convergent evolution.*

Carnitine carries long-chain activated fatty acids into the mitochondrial matrix

Fatty acids are activated on the outer mitochondrial membrane, whereas they are oxidized in the mitochondrial matrix. A special transport mechanism is needed to carry activated long-chain fatty acids across the inner mitochondrial membrane. These fatty acids must be conjugated to *carnitine,* an alcohol with both a positive and a negative charge (a zwitterion). The acyl group is transferred from the sulfur atom of coenzyme A to the hydroxyl group of carnitine to form *acyl carnitine*. This reaction is catalyzed by *carnitine acyltransferase I,* also called *carnitine palmitoyl transferase I,* which is bound to the outer mitochondrial membrane.



Acyl carnitine is then shuttled across the inner mitochondrial membrane by a translocase (Figure 22.8). The acyl group is transferred back to coenzyme A on the matrix side of the membrane. This reaction, which is catalyzed by *carnitine acyltransferase II (carnitine palmitoyl transferase II),* is simply the reverse of the reaction that takes place in the cytoplasm. The reaction is thermodynamically feasible because of the zwitterionic nature of carnitine. The *O*-acyl link in carnitine has a high group-transfer potential, apparently because, being zwitterions, carnitine and its esters are solvated differently from most other alcohols and their esters. Finally, the translocase returns carnitine to the cytoplasmic side in exchange for an incoming acyl carnitine.

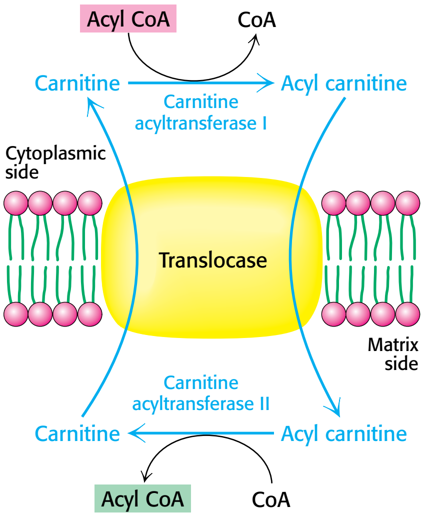
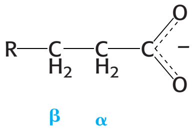


FIGURE 22.8 Acyl carnitine translocase. The entry of acyl carnitine into the mitochondrial matrix is mediated by a translocase. Carnitine returns to the cytoplasmic side of the inner mitochondrial membrane in exchange for acyl carnitine.

A number of diseases have been traced to a deficiency of carnitine, the transferase, or the translocase. Inability to synthesize carnitine may be a contributing factor to the development of autism in males. The symptoms of carnitine deficiency range from mild muscle cramping to severe weakness and even death. In general, muscle, kidney, and heart are the tissues primarily impaired. Muscle weakness during prolonged exercise is a symptom of a deficiency of carnitine acyltransferases because muscle relies on fatty acids as a long-term source of energy. Medium-chain (C_8–C_{10}) fatty acids are oxidized normally in these patients because these fatty acids can enter the mitochondria, to some degree, in the absence of carnitine. These diseases illustrate that *the impaired flow of a metabolite from one compartment of a cell to another can lead to a pathological condition.*

Acetyl CoA, NADH, and FADH$_2$ are generated in each round of fatty acid oxidation

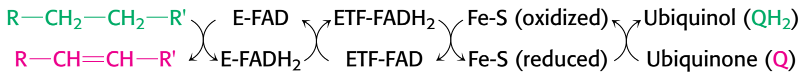
A saturated acyl CoA is degraded by a recurring sequence of four reactions: oxidation by flavin adenine dinucleotide (FAD), hydration, oxidation by NAD$^+$, and thiolysis by coenzyme A (Figure 22.9). The fatty acid chain is shortened by two carbon atoms as a result of these reactions and FADH$_2$, NADH, and acetyl CoA are generated. Because oxidation takes place at the β carbon atom, this series of reactions is called the *β-oxidation pathway.*



The first reaction in each round of degradation is the *oxidation* of acyl CoA by an *acyl CoA dehydrogenase* to give an enoyl CoA with a *trans* double bond between C-2 and C-3.

$$\text{Acyl CoA} + \text{E-FAD} \longrightarrow \textit{trans}\text{-}\Delta^2\text{-enoyl CoA} + \text{E-FADH}_2$$

As in the dehydrogenation of succinate in the citric acid cycle, FAD rather than NAD$^+$ is the electron acceptor because the ΔG for this reaction is insufficient to drive the reduction of NAD^+. Electrons from the $FADH_2$ prosthetic group of the reduced acyl CoA dehydrogenase are transferred to a second flavoprotein called *electron-transferring flavoprotein (ETF).* In turn, ETF donates electrons to *ETF: ubiquinone reductase,* an iron–sulfur protein. Ubiquinone is thereby reduced to ubiquinol, which delivers its high-potential electrons to the second proton-pumping site of the respiratory chain (Section 18.3). Consequently, 1.5 molecules of ATP are generated per molecule of $FADH_2$ formed in this dehydrogenation step, as in the oxidation of succinate to fumarate.



The next step is the *hydration* of the double bond between C-2 and C-3 by *enoyl CoA hydratase.*

$$trans\text{-}\Delta^2\text{-Enoyl CoA} + H_2O \longrightarrow \text{L-3-hydroxyacyl CoA}$$

The hydration of enoyl CoA is stereospecific. Only the L isomer of 3-hydroxyacyl CoA is formed when the *trans*-Δ^2 double bond is hydrated. The enzyme also hydrates a *cis*-Δ^2 double bond, but the product then is the D isomer. We shall return to this point shortly in considering how unsaturated fatty acids are oxidized.

The hydration of enoyl CoA is a prelude to the second *oxidation* reaction, which converts the hydroxyl group at C-3 into a keto group and generates NADH. This oxidation is catalyzed by *L-3-hydroxyacyl CoA dehydrogenase,* which is specific for the L isomer of the hydroxyacyl substrate.

$$\text{L-3-Hydroxyacyl CoA} + NAD^+ \rightleftharpoons \text{3-ketoacyl CoA} + NADH + H^+$$

The preceding reactions have oxidized the methylene group at C-3 to a keto group. The final step is the *cleavage* of 3-ketoacyl CoA by the thiol group of a second molecule of coenzyme A, which yields acetyl CoA and an acyl CoA shortened by two carbon atoms. This thiolytic cleavage is catalyzed by *β-ketothiolase.*

$$\underset{(n \text{ carbons})}{\text{3-Ketoacyl CoA}} + \text{HS-CoA} \rightleftharpoons \text{acetyl CoA} + \underset{(n-2 \text{ carbons})}{\text{acyl CoA}}$$

Table 22.1 summarizes the reactions in fatty acid oxidation.

The shortened acyl CoA then undergoes another cycle of oxidation, starting with the reaction catalyzed by acyl CoA dehydrogenase

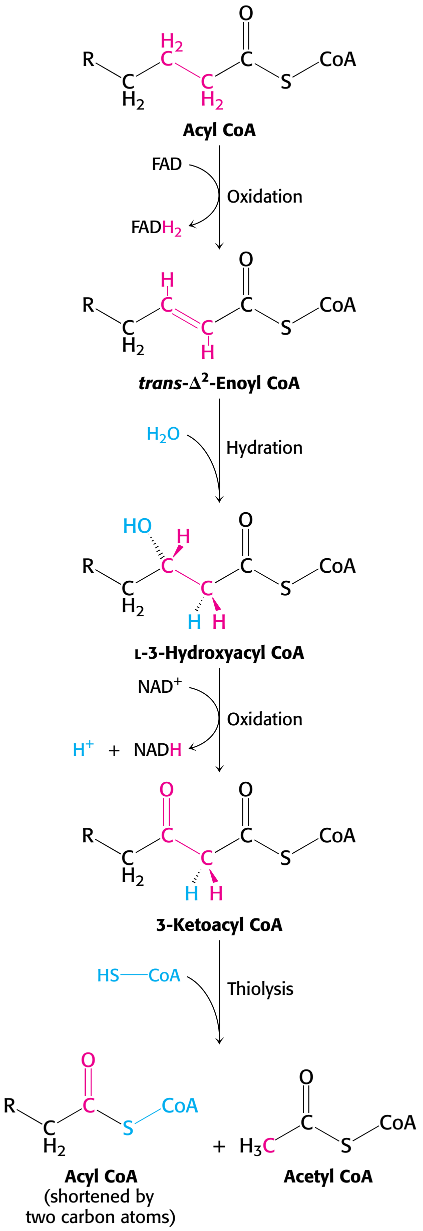


FIGURE 22.9 Reaction sequence for the degradation of fatty acids. Fatty acids are degraded by the repetition of a four-reaction sequence consisting of oxidation, hydration, oxidation, and thiolysis.

TABLE 22.1 Principal reactions in fatty acid oxidation

Step	Reaction	Enzyme
1	Fatty acid + CoA + ATP $\rightleftharpoons$ acyl CoA + AMP + PP_i	Acyl CoA synthetase (also called fatty acid thiokinase and fatty acid:CoA ligase)*
2	Carnitine + acyl CoA $\rightleftharpoons$ acyl carnitine + CoA	Carnitine acyltransferase (also called carnitine palmitoyl transferase)
3	Acyl CoA + E-FAD ⟶ *trans*-Δ^2-enoyl CoA + E-$FADH_2$	Acyl CoA dehydrogenases (several isozymes having different chain-length specificity)
4	*trans*-Δ^2-Enoyl CoA + H_2O $\rightleftharpoons$ L-3-hydroxyacyl CoA	Enoyl CoA hydratase (also called crotonase or 3-hydroxyacyl CoA hydrolyase)
5	L-3-Hydroxyacyl CoA + NAD^+ $\rightleftharpoons$ 3-ketoacyl CoA + NADH + H^+	L-3-Hydroxyacyl CoA dehydrogenase
6	3-Ketoacyl CoA + CoA $\rightleftharpoons$ acetyl CoA + acyl CoA (shortened by C_2)	β-Ketothiolase (also called thiolase)

*An AMP-forming ligase.

(Figure 22.10). Fatty acid chains containing from 12 to 18 carbon atoms are oxidized by the long-chain acyl CoA dehydrogenase. The medium-chain acyl CoA dehydrogenase oxidizes fatty acid chains having from 14 to 4 carbons, whereas the short-chain acyl CoA dehydrogenase acts only on 4- and 6-carbon fatty acid chains. In contrast, β-ketothiolase, hydroxyacyl dehydrogenase, and enoyl CoA hydratase act on fatty acid molecules of almost any length.

The complete oxidation of palmitate yields 106 molecules of ATP

We can now calculate the energy yield derived from the oxidation of a fatty acid. In each reaction cycle, an acyl CoA is shortened by two carbon atoms, and one molecule each of $FADH_2$, NADH, and acetyl CoA are formed.

$$\begin{aligned}&C_n\text{-acyl CoA} + \text{FAD} + \text{NAD}^+ + H_2O + \text{CoA} \longrightarrow \\ &\qquad C_{n-2}\text{-acyl CoA} + FADH_2 + \text{NADH} + \text{acetyl CoA} + H^+\end{aligned}$$

The degradation of palmitoyl CoA (C_{16}-acyl CoA) requires seven reaction cycles. In the seventh cycle, the C_4-ketoacyl CoA is thiolyzed to two molecules of acetyl CoA. Hence, the stoichiometry of the oxidation of palmitoyl CoA is

$$\begin{aligned}&\text{Palmitoyl CoA} + 7\ \text{FAD} + 7\ \text{NAD}^+ + 7\ \text{CoA} + 7\ H_2O \longrightarrow \\ &\qquad 8\ \text{acetyl CoA} + 7\ FADH_2 + 7\ \text{NADH} + 7\ H^+\end{aligned}$$

Approximately 2.5 molecules of ATP are generated when the respiratory chain oxidizes each of these NADH molecules, whereas 1.5 molecules of ATP are formed for each $FADH_2$ because their electrons enter the chain at the level of ubiquinol. Recall that the oxidation of acetyl CoA by the citric acid cycle yields 10 molecules of ATP. Hence, the number of ATP molecules formed in the oxidation of palmitoyl CoA is 10.5 from the seven $FADH_2$, 17.5 from the seven NADH, and 80 from the eight acetyl CoA molecules, which gives a total of 108. The equivalent of 2 molecules of ATP is consumed in the activation of palmitate, in which ATP is split into AMP and 2 molecules of orthophosphate. Thus, *the complete oxidation of a molecule of palmitate yields 106 molecules of ATP.*

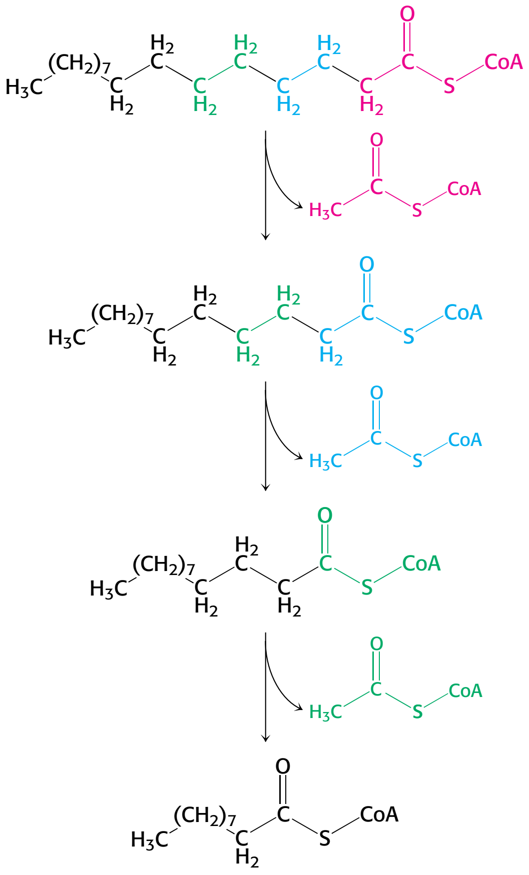


FIGURE 22.10 First three rounds in the degradation of palmitate. Two-carbon units are sequentially removed from the carboxyl end of the fatty acid.

22.3 Unsaturated and Odd-Chain Fatty Acids Require Additional Steps for Degradation

The β-oxidation pathway accomplishes the complete degradation of saturated fatty acids having an even number of carbon atoms. Most fatty acids have such structures because of their mode of synthesis (to be addressed later in this chapter). However, not all fatty acids are so simple. The oxidation of fatty acids containing double bonds requires additional steps, as does the oxidation of fatty acids containing an odd number of carbon atoms.

An isomerase and a reductase are required for the oxidation of unsaturated fatty acids

The oxidation of unsaturated fatty acids presents some difficulties, yet many such fatty acids are available in the diet. Most of the reactions are the same as those for saturated fatty acids. In fact, only two additional enzymes—an isomerase and a reductase—are needed to degrade a wide range of unsaturated fatty acids.

Consider the oxidation of palmitoleate (Figure 22.11). This C_{16} unsaturated fatty acid, which has one double bond between C-9 and C-10,

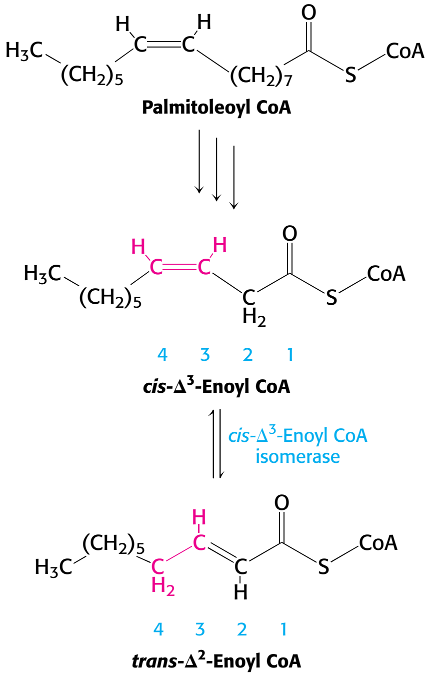


FIGURE 22.11 The degradation of a monounsaturated fatty acid. *Cis*-Δ^3-Enoyl CoA isomerase allows continued β-oxidation of fatty acids with a single double bond.

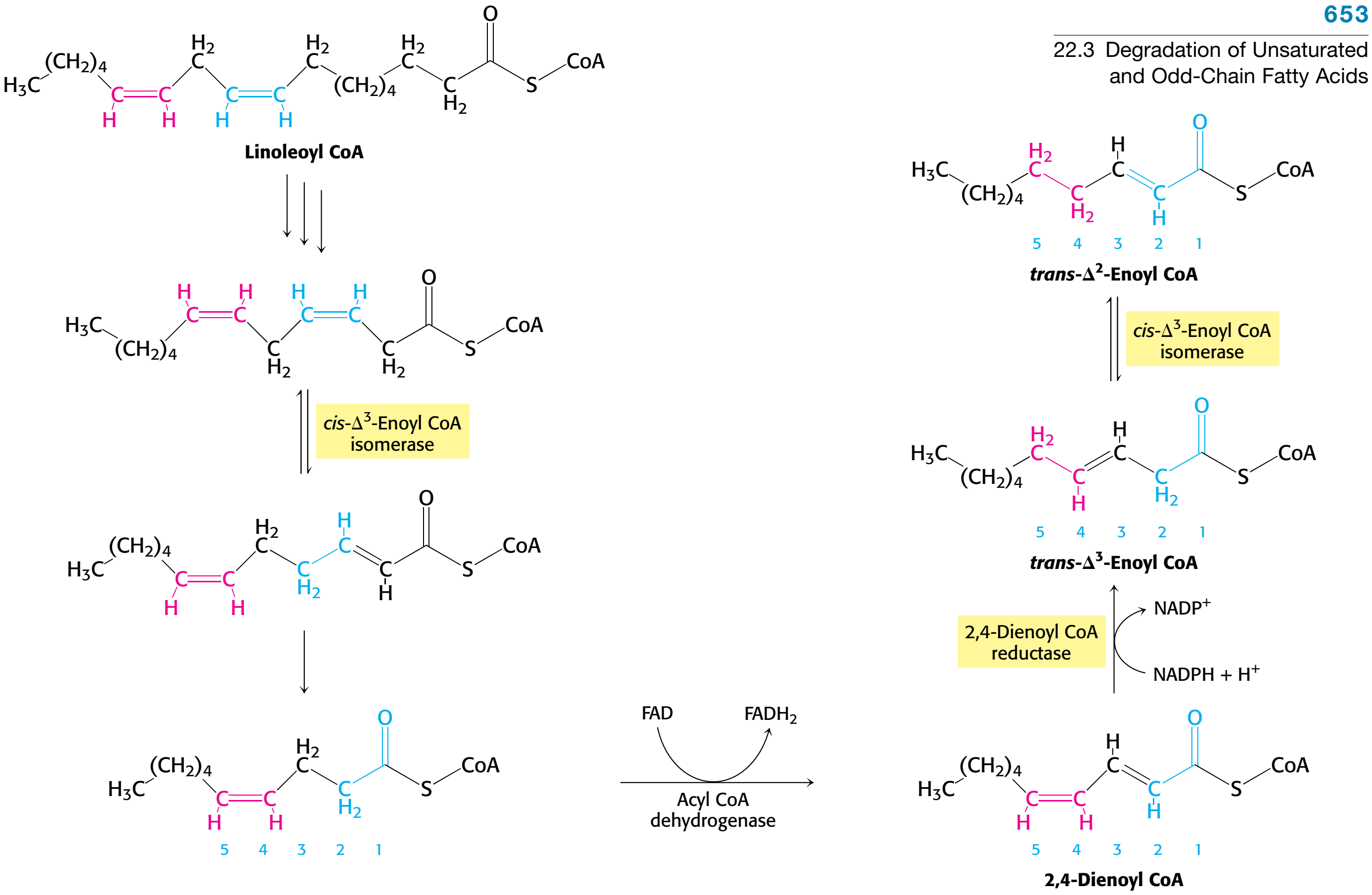


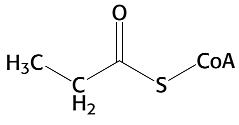
FIGURE 22.12 Oxidation of linoleoyl CoA. The complete oxidation of the diunsaturated fatty acid linoleate is facilitated by the activity of enoyl CoA isomerase and 2,4-dienoyl CoA reductase.

is activated and transported across the inner mitochondrial membrane in the same way as saturated fatty acids are. Palmitoleoyl CoA then undergoes three cycles of degradation, which are carried out by the same enzymes as those in the oxidation of saturated fatty acids. However, the *cis*-Δ^3-enoyl CoA formed in the third round is not a substrate for acyl CoA dehydrogenase. The presence of a double bond between C-3 and C-4 prevents the formation of another double bond between C-2 and C-3. This impasse is resolved by a new reaction that shifts the position and configuration of the *cis*-Δ^3 double bond. *cis-*Δ^3*-Enoyl CoA isomerase converts this double bond into a trans-*Δ^2 *double bond.* The double bond is now between C-2 and C-3. The subsequent reactions are those of the saturated fatty acid oxidation pathway, in which the *trans*-Δ^2-enoyl CoA is a regular substrate.

Human beings require polyunsaturated fatty acids, which have multiple double bonds, as important precursors for signal molecules, but excess polyunsaturated fatty acids are degraded by β oxidation. However, another problem arises with the oxidation of polyunsaturated fatty acids. Consider linoleate, a C_{18} polyunsaturated fatty acid with *cis*-Δ^9 and *cis*-Δ^{12} double bonds (Figure 22.12). The *cis*-Δ^3 double bond (between carbons 3 and 4) formed after three rounds of β-oxidation is converted into a *trans*-Δ^2 double bond (between carbons 2 and 3) by the aforementioned isomerase. The acyl CoA produced by another round of β-oxidation contains a *cis*-Δ^4 (between carbons 4 and 5) double bond. Dehydrogenation of this species by acyl CoA dehydrogenase yields a *2,4-dienoyl intermediate* (double bond between carbons 2 and 3 and carbons 4 and 5), which is not a substrate for the next enzyme in the β-oxidation pathway. This impasse is circumvented by *2,4-dienoyl CoA reductase,* an enzyme that uses

NADPH to reduce the 2,4-dienoyl intermediate to *trans*-Δ^3-enoyl CoA. *cis*-Δ^3-Enoyl CoA isomerase then converts *trans*-Δ^3-enoyl CoA into the *trans*-Δ^2 form, a customary intermediate in the β-oxidation pathway. These catalytic strategies are elegant and economical. Only two extra enzymes are needed for the oxidation of *any* polyunsaturated fatty acid. *Odd-numbered double bonds are handled by the isomerase, and even-numbered ones by the reductase and the isomerase.*

Odd-chain fatty acids yield propionyl CoA in the final thiolysis step



Propionyl CoA

Fatty acids having an odd number of carbon atoms are minor species. They are oxidized in the same way as fatty acids having an even number, except that propionyl CoA and acetyl CoA, rather than two molecules of acetyl CoA, are produced in the final round of degradation. The activated three-carbon unit in propionyl CoA enters the citric acid cycle after it has been converted into succinyl CoA.

The pathway from propionyl CoA to succinyl CoA is especially interesting because it entails a rearrangement that requires *vitamin* B_{12} (also known as *cobalamin*). Propionyl CoA is carboxylated at the expense of the hydrolysis of a molecule of ATP to yield the D isomer of methylmalonyl CoA (Figure 22.13). This carboxylation reaction is catalyzed by *propionyl CoA carboxylase,* a biotin enzyme that has a catalytic mechanism like that of the homologous enzyme pyruvate carboxylase (Section 16.3). The D isomer of methylmalonyl CoA is racemized to the L isomer, the substrate for a mutase that converts it into *succinyl CoA* by an *intramolecular rearrangement.* The —CO—S—CoA group migrates from C-2 to a methyl group in exchange for a hydrogen atom. This very unusual isomerization is catalyzed by *methylmalonyl CoA mutase,* which contains a derivative of cobalamin as its coenzyme.

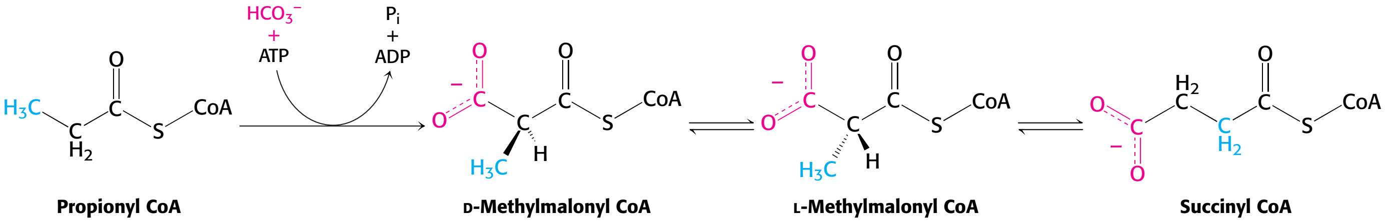


FIGURE 22.13 Conversion of propionyl CoA into succinyl CoA. Propionyl CoA, generated from fatty acids with an odd number of carbons as well as some amino acids, is converted into the citric acid cycle intermediate succinyl CoA.

Vitamin B_{12} contains a corrin ring and a cobalt atom

Cobalamin enzymes, which are present in most organisms, catalyze three types of reactions: (1) *intramolecular rearrangements;* (2) *methylations,* as in the synthesis of methionine; and (3) the *reduction of ribonucleotides to deoxyribonucleotides* (Section 25.3). In mammals, only two reactions are known to require coenzyme B_{12}. The conversion of L-methylmalonyl CoA into succinyl CoA is one, and the formation of methionine by the methylation of homocysteine is the other (Section 24.2). The latter reaction is especially important because methionine is required for the generation of coenzymes that participate in the synthesis of purines and thymine, which are needed for nucleic acid synthesis.

The core of cobalamin consists of a *corrin ring with a central cobalt atom* (Figure 22.14). The corrin ring, like a porphyrin, has *four pyrrole units.* Two of them are directly bonded to each other, whereas the others are joined by methine bridges, as in porphyrins. The corrin ring is more reduced than that of porphyrins and the substituents are different. A cobalt atom is bonded to the four pyrrole nitrogens. The fifth substituent linked to the cobalt atom is

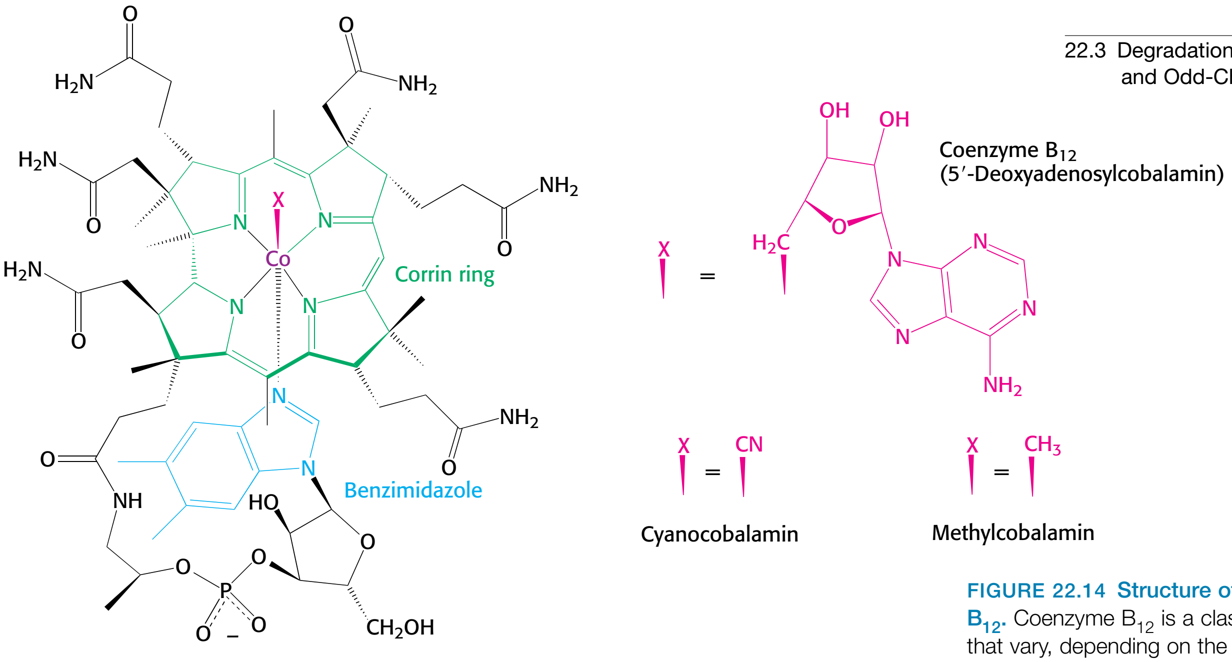


FIGURE 22.14 Structure of coenzyme B_{12}. Coenzyme B_{12} is a class of molecules that vary, depending on the component designated X in the left-hand structure. 5′-Deoxyadenosylcobalamin is the form of the coenzyme in methylmalonyl mutase. Substitution of methyl and cyano groups for X creates methylcobalamin and, cyanocobalamin, respectively.

a derivative of *dimethylbenzimidazole* that contains ribose 3-phosphate and aminoisopropanol. One of the nitrogen atoms of dimethylbenzimidazole is linked to the cobalt atom. In coenzyme B_{12}, *the sixth substituent* linked to the cobalt atom is a *5′-deoxyadenosyl unit* or a methyl group. This position can also be occupied by a cyano group. Cyanocobalamin is the form of the coenzyme administered to treat B_{12} deficiency. In all of these compounds, the cobalt is in the +3 oxidation state.

Mechanism: Methylmalonyl CoA mutase catalyzes a rearrangement to form succinyl CoA

The rearrangement reactions catalyzed by coenzyme B_{12} are exchanges of two groups attached to adjacent carbon atoms of the substrate (Figure 22.15). A hydrogen atom migrates from one carbon atom to the next, and an R group (such as the —CO—S—CoA group of methylmalonyl CoA) concomitantly moves in the reverse direction. The first step in these intramolecular rearrangements is the cleavage of the carbon–cobalt bond of 5′-deoxyadenosylcobalamin to generate the Co^{2+} form of the coenzyme and a 5′-deoxyadenosyl radical, $—CH_2·$ (Figure 22.16). In this *homolytic cleavage reaction,* one electron of the Co–C bond stays with Co (reducing it from the +3 to the +2 oxidation state), whereas the other electron stays with the carbon atom, generating a free radical. In contrast, nearly all other cleavage reactions in biological systems are *heterolytic:* an electron *pair* is transferred to one of the two atoms that were bonded together.

What is the role of this very unusual $—CH_2·$ radical? This highly reactive species abstracts a *hydrogen atom* from the substrate to form

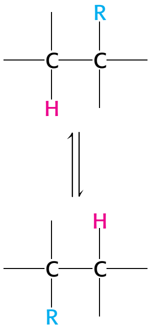


FIGURE 22.15 Rearrangement reaction catalyzed by cobalamin enzymes. The R group can be an amino group, a hydroxyl group, or a substituted carbon.

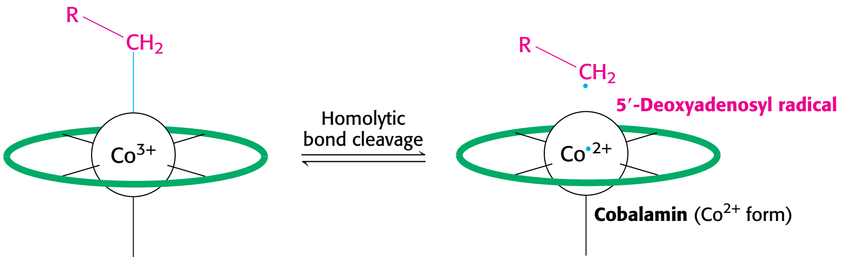


FIGURE 22.16 Formation of a 5′-deoxyadenosyl radical. The methylmalonyl CoA mutase reaction begins with the homolytic cleavage of the bond joining Co^{3+} of coenzyme B_{12} to a carbon atom of the ribose of the adenosine moiety of the enzyme. The cleavage generates a 5′-deoxyadenosyl radical and leads to the reduction of Co^{3+} to Co^{2+}. The letter R represents the 5′-deoxyadenosyl component of the coenzyme, and the green oval represents the remainder of the coenzyme.

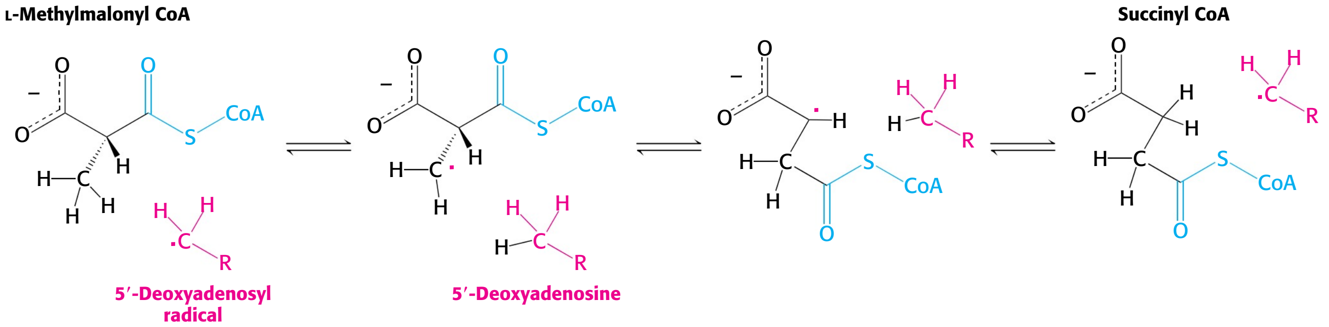

L-Methylmalonyl CoA
Succinyl CoA
CoA
5′-Deoxyadenosyl radical
5′-Deoxyadenosine

FIGURE 22.17 Formation of succinyl CoA by a rearrangement reaction. A free radical abstracts a hydrogen atom in the rearrangement of methylmalonyl CoA to succinyl CoA.

5′-deoxyadenosine and a substrate radical (Figure 22.17). This substrate radical spontaneously rearranges: the carbonyl CoA group migrates to the position formerly occupied by H on the neighboring carbon atom to produce a different radical. This product radical abstracts a hydrogen atom from the methyl group of 5′-deoxyadenosine to complete the rearrangement and return the deoxyadenosyl unit to the radical form. *The role of coenzyme B_{12} in such intramolecular migrations is to serve as a source of free radicals for the abstraction of hydrogen atoms.*

An essential property of coenzyme B_{12} is the weakness of its cobalt–carbon bond, which is readily cleaved to generate a radical. To facilitate the cleavage of this bond, enzymes such as methylmalonyl CoA mutase displace the benzimidazole group from the cobalamin and bind to the cobalt atom through a histidine residue (Figure 22.18). The steric crowding around the cobalt–carbon bond within the corrin ring system contributes to the bond weakness.

Fatty acids are also oxidized in peroxisomes

Although most fatty acid oxidation takes place in mitochondria, oxidation of long chain and branched fatty acids takes place in cellular organelles

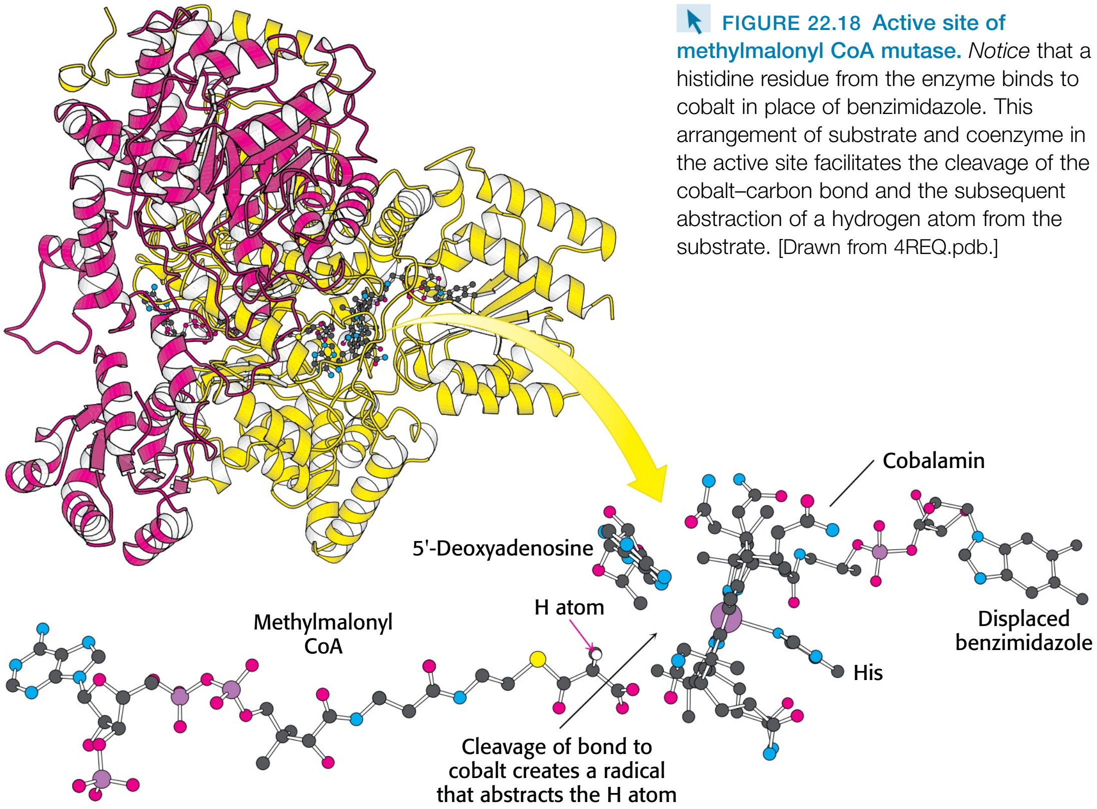

Cobalamin
5'-Deoxyadenosine
Methylmalonyl CoA
H atom
Displaced benzimidazole
His
Cleavage of bond to cobalt creates a radical that abstracts the H atom

FIGURE 22.18 Active site of methylmalonyl CoA mutase. *Notice* that a histidine residue from the enzyme binds to cobalt in place of benzimidazole. This arrangement of substrate and coenzyme in the active site facilitates the cleavage of the cobalt–carbon bond and the subsequent abstraction of a hydrogen atom from the substrate. [Drawn from 4REQ.pdb.]

called *peroxisomes* (Figure 22.19) (problems 17 and 44). These organelles are small membrane-bounded compartments that are present in the cells of most eukaryotes. Fatty acid oxidation in these organelles, which halts at octanoyl CoA, serves to shorten very long chains (C_{26}) to make them better substrates of β oxidation in mitochondria. Peroxisomal oxidation differs from β oxidation in the initial dehydrogenation reaction (Figure 22.20). In peroxisomes, acyl CoA dehydrogenase, a flavoprotein, transfers electrons from the substrate to $FADH_2$ and then to O_2 to yield H_2O_2 instead of capturing high-energy electrons as $FADH_2$ for use in the electron-transport chain, as in mitochondrial β oxidation. Peroxisomes contain high concentrations of the enzyme catalase to degrade H_2O_2 into water and O_2. Subsequent steps are identical with those of their mitochondrial counterparts, although they are carried out by different isoforms of the enzymes.

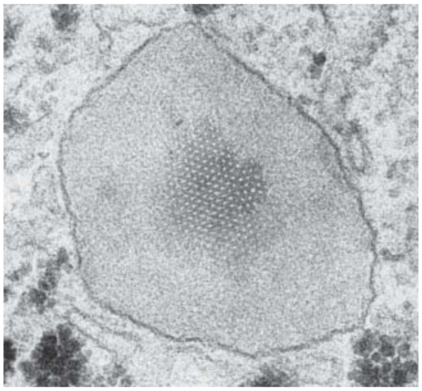


FIGURE 22.19 Electron micrograph of a peroxisome in a liver cell. A crystal of urate oxidase is present inside the organelle, which is bounded by a single bilayer membrane. The dark granular structures outside the peroxisome are glycogen particles. [Courtesy of Dr. George Palade.]

Peroxisomes do not function in patients with Zellweger syndrome. Liver, kidney, and muscle abnormalities usually lead to death by age 6. The syndrome is caused by a defect in the import of enzymes into the peroxisomes. Here we see a pathological condition resulting from an inappropriate cellular distribution of enzymes.

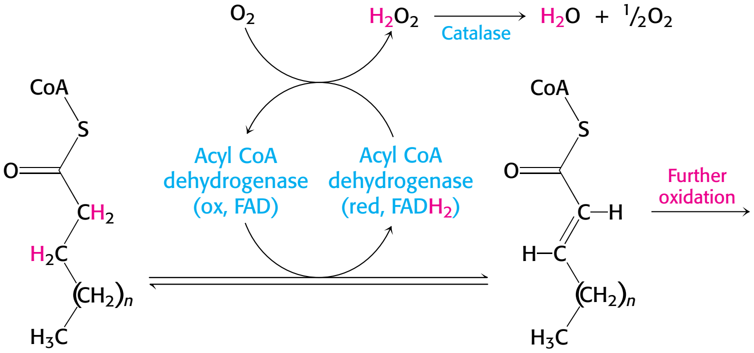


FIGURE 22.20 Initiation of peroxisomal fatty acid degradation. The first dehydrogenation in the degradation of fatty acids in peroxisomes requires a flavoprotein dehydrogenase that transfers electrons from its $FADH_2$ moiety to O_2 to yield H_2O_2.

Ketone bodies are formed from acetyl CoA when fat breakdown predominates

The acetyl CoA formed in fatty acid oxidation enters the citric acid cycle only if fat and carbohydrate degradation are appropriately balanced. Acetyl CoA must combine with oxaloacetate to gain entry to the citric acid cycle. The availability of oxaloacetate, however, depends on an adequate supply of carbohydrate. Recall that oxaloacetate is normally formed from pyruvate, the product of glucose degradation in glycolysis, by pyruvate carboxylase (Section 16.3). If carbohydrate is unavailable or improperly utilized, the concentration of oxaloacetate is lowered and acetyl CoA cannot enter the citric acid cycle. This dependency is the molecular basis of the adage that *fats burn in the flame of carbohydrates.*

In fasting or diabetes, oxaloacetate is consumed to form glucose by the gluconeogenic pathway (Section 16.3) and hence is unavailable for condensation with acetyl CoA. Under these conditions, acetyl CoA is diverted to the formation of acetoacetate and D-3-hydroxybutyrate. Acetoacetate, D-3-hydroxybutyrate, and acetone are often referred to as *ketone bodies.* Abnormally high levels of ketone bodies are present in the blood of untreated diabetics.

Acetoacetate is formed from acetyl CoA in three steps (Figure 22.21). Two molecules of acetyl CoA condense to form acetoacetyl CoA. This reaction, which is catalyzed by thiolase, is the reverse of the thiolysis step in the oxidation of fatty acids. Acetoacetyl CoA then reacts with acetyl

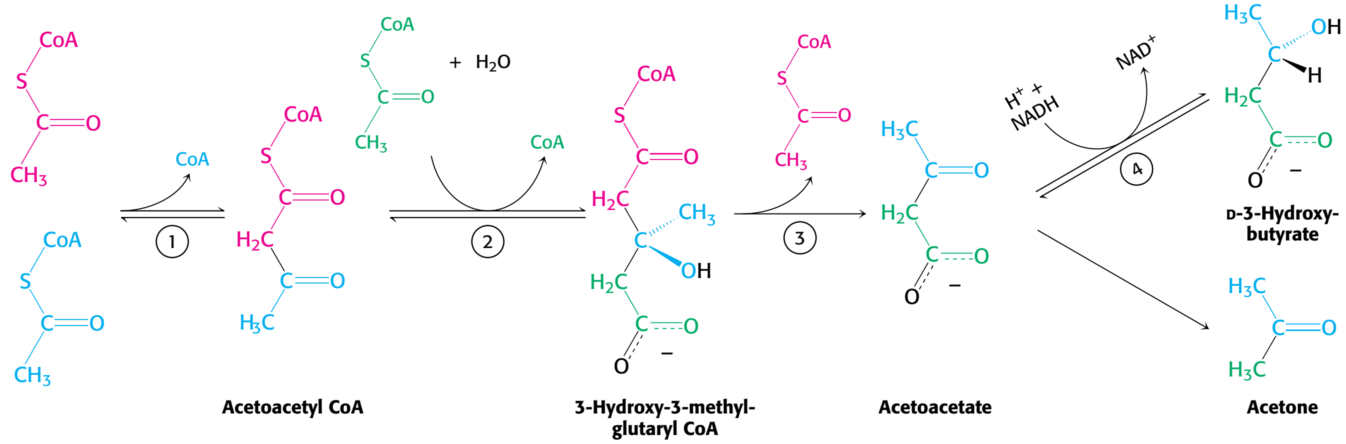


FIGURE 22.21 Formation of ketone bodies. The ketone bodies—acetoacetate, D-3-hydroxybutyrate, and acetone—are formed from acetyl CoA primarily in the liver. Enzymes catalyzing these reactions are (1) 3-ketothiolase, (2) hydroxymethylglutaryl CoA synthase, (3) hydroxymethylglutaryl CoA cleavage enzyme, and (4) D-3-hydroxybutyrate dehydrogenase. Acetoacetate spontaneously decarboxylates to form acetone.

CoA and water to give 3-hydroxy-3-methylglutaryl CoA (HMG-CoA) and CoA. This condensation resembles the one catalyzed by citrate synthase (Section 17.2). This reaction, which has a favorable equilibrium owing to the hydrolysis of a thioester linkage, compensates for the unfavorable equilibrium in the formation of acetoacetyl CoA. 3-Hydroxy-3-methylglutaryl CoA is then cleaved to acetyl CoA and acetoacetate. The sum of these reactions is

$$2\ \text{Acetyl CoA} + H_2O \longrightarrow \text{acetoacetate} + 2\ \text{CoA} + H^+$$

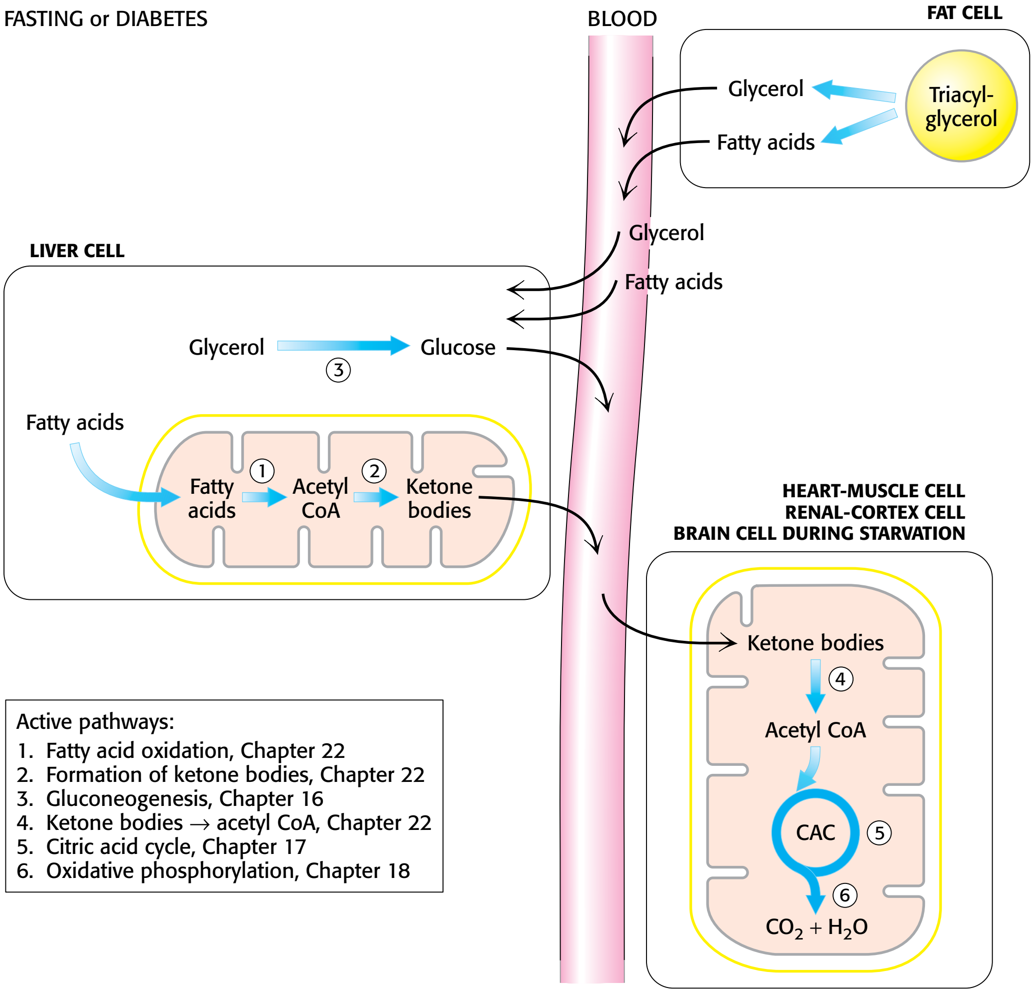
D-3-Hydroxybutyrate is formed by the reduction of acetoacetate in the mitochondrial matrix by D-3-hydroxybutyrate dehydrogenase. The ratio of hydroxybutyrate to acetoacetate depends on the $NADH/NAD^+$ ratio inside mitochondria.

Because it is a β-ketoacid, acetoacetate also undergoes a slow, spontaneous decarboxylation to acetone. The odor of acetone may be detected in the breath of a person who has a high level of acetoacetate in the blood. Under starvation conditions, the acetone may be captured to synthesize glucose.

Ketone bodies are a major fuel in some tissues

The major site of the production of acetoacetate and 3-hydroxybutyrate is the liver. These molecules diffuse from the liver mitochondria into the blood and are transported to other tissues such as heart and kidney (Figure 22.22). Acetoacetate and 3-hydroxybutyrate are normal fuels of respiration and are quantitatively important as sources of energy. Indeed, heart muscle and the renal cortex use acetoacetate in preference to glucose. In contrast, glucose is the major fuel for the brain and red blood cells in well-nourished people on a balanced diet. However, the brain adapts to the utilization of acetoacetate during starvation and diabetes. In prolonged starvation, 75% of the fuel needs of the brain are met by ketone bodies.

Acetoacetate is converted into acetyl CoA in two steps. First, acetoacetate is activated by the transfer of CoA from succinyl CoA in a reaction catalyzed by a specific CoA transferase. Second, acetoacetyl CoA is cleaved by thiolase to yield two molecules of acetyl CoA, which can then enter the citric acid


FASTING or DIABETES
BLOOD
FAT CELL
Glycerol
Fatty acids
Triacyl-glycerol
Glycerol
Fatty acids
LIVER CELL
Glycerol
③
Glucose
Fatty acids
Fatty acids
①
Acetyl CoA
②
Ketone bodies
HEART-MUSCLE CELL
RENAL-CORTEX CELL
BRAIN CELL DURING STARVATION
Ketone bodies
④
Acetyl CoA
CAC
⑤
⑥
$CO_2 + H_2O$

Active pathways:
1. Fatty acid oxidation, Chapter 22
2. Formation of ketone bodies, Chapter 22
3. Gluconeogenesis, Chapter 16
4. Ketone bodies → acetyl CoA, Chapter 22
5. Citric acid cycle, Chapter 17
6. Oxidative phosphorylation, Chapter 18

FIGURE 22.22 PATHWAY INTEGRATION: Liver supplies ketone bodies to the peripheral tissues. During fasting or in untreated diabetics, the liver converts fatty acids into ketone bodies, which are a fuel source for a number of tissues. Ketone-body production is especially important during starvation, when ketone bodies are the predominant fuel.

cycle (Figure 22.23). The liver has acetoacetate available to supply to other organs because it lacks this particular CoA transferase. 3-Hydroxybutyrate requires an additional step to yield acetyl CoA. It is first oxidized to produce acetoacetate, which is processed as heretofore described, and NADH for use in oxidative phosphorylation.

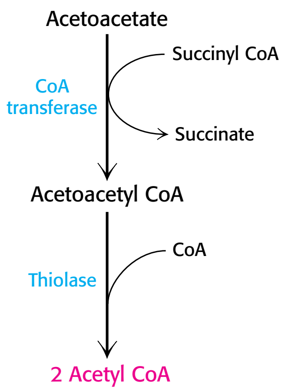
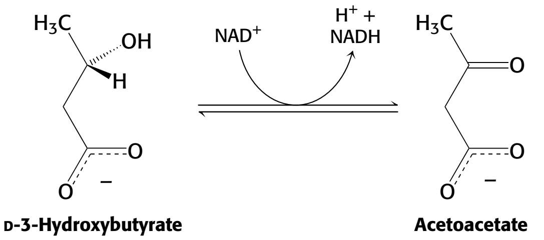

Acetoacetate
Succinyl CoA
CoA transferase
Succinate
Acetoacetyl CoA
CoA
Thiolase
2 Acetyl CoA

FIGURE 22.23 Utilization of acetoacetate as a fuel. Acetoacetate can be converted into two molecules of acetyl CoA, which then enter the citric acid cycle.


NAD^+
$H^+ + NADH$
D-3-Hydroxybutyrate
Acetoacetate

Ketone bodies can be regarded as a water-soluble, transportable form of acetyl units. Fatty acids are released by adipose tissue and converted into acetyl units by the liver, which then exports them as acetoacetate. As might be expected, acetoacetate also has a regulatory role. *High levels of acetoacetate in the blood signify an abundance of acetyl units and lead to a decrease in the rate of lipolysis in adipose tissue.*

High blood levels of ketone bodies, the result of certain pathological conditions, can be life threatening. The most common of these conditions is diabetic ketosis in patients with insulin-dependent diabetes

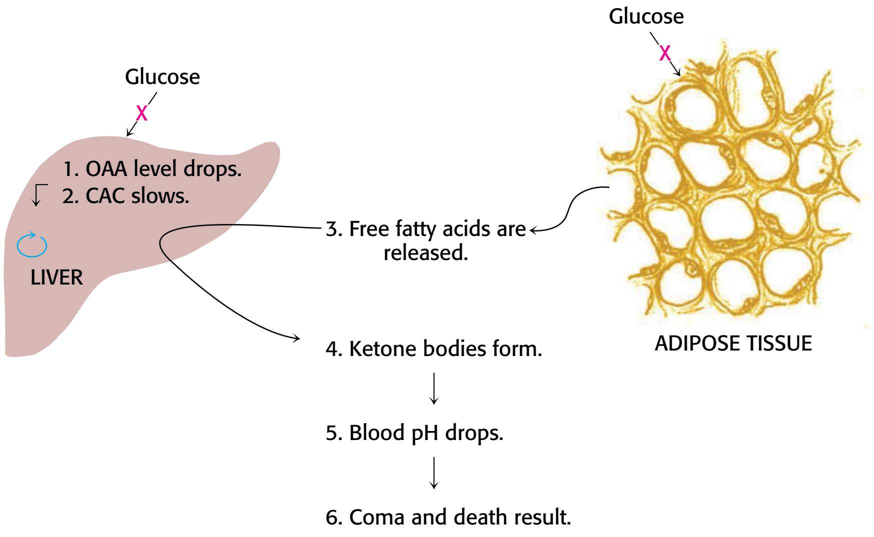


FIGURE 22.24 Diabetic ketosis results when insulin is absent. In the absence of insulin, fats are released from adipose tissue, and glucose cannot be absorbed by the liver or adipose tissue. The liver degrades the fatty acids by β oxidation but cannot process the acetyl CoA, because of a lack of glucose-derived oxaloacetate (OAA). Excess ketone bodies are formed and released into the blood.

mellitus. These patients are unable to produce insulin. As stated earlier, this hormone, normally released after meals, signals tissues to take up glucose. In addition, it curtails fatty acid mobilization by adipose tissue. The absence of insulin has two major biochemical consequences (Figure 22.24). First, the liver cannot absorb glucose and consequently cannot provide oxaloacetate to process fatty acid-derived acetyl CoA. Second, adipose cells continue to release fatty acids into the blood stream; these fatty acids are then taken up by the liver and converted into ketone bodies. The liver thus produces large amounts of ketone bodies, which are moderately strong acids. The result is severe acidosis. The decrease in pH impairs tissue function, most importantly in the central nervous system.

Interestingly, diets that promote ketone-body formation, called *ketogenic diets,* are frequently used as a therapeutic option for children with drug-resistant epilepsy. Ketogenic diets are rich in fats and low in carbohydrates, with adequate amounts of protein. In essence, the body is forced into starvation mode, where fats and ketone bodies become the main fuel source (Section 27.5). How such diets reduce the seizures suffered by the children is currently unknown.

Animals cannot convert fatty acids into glucose

A typical human being has far greater fat stores than glycogen stores. However, glycogen is necessary to fuel very active muscle, as well as the brain, which normally uses only glucose as a fuel. When glycogen stores are low, why can't the body make use of fat stores and convert fatty acids into glucose? Because *animals are unable to effect the net synthesis of glucose from fatty acids.* Specifically, acetyl CoA cannot be converted into pyruvate or oxaloacetate in animals. Recall that the reaction that generates acetyl CoA from pyruvate is irreversible (Section 17.1). The two carbon atoms of the acetyl group of acetyl CoA enter the citric acid cycle, but two carbon atoms leave the cycle in the decarboxylations catalyzed by isocitrate dehydrogenase and α-ketoglutarate dehydrogenase. Consequently, oxaloacetate is regenerated, but it is not formed de novo when the acetyl unit of acetyl CoA is oxidized by the citric acid cycle. In essence, two carbon atoms enter the cycle as an acetyl group, but two carbons leave the cycle as CO_2 before oxaloacetate is generated. As a result, no net synthesis of oxaloacetate is possible. In contrast, plants have two additional enzymes enabling them to convert the carbon atoms of acetyl CoA into oxaloacetate (Section 17.5).

Some fatty acids may contribute to the development of pathological conditions

As we will see shortly (Section 22.5), certain polyunsaturated fatty acids are essential for life, serving as precursors to various signal molecules. Vegetable oils, used commonly in food preparation, are rich in polyunsaturated fatty acids. However, polyunsaturated fatty acids are unstable and are readily oxidized. This tendency to become rancid reduces their shelf life and renders them undesirable for cooking. To circumvent this problem, polyunsaturated fatty acids are hydrogenated, converting them to saturated and trans unsaturated fatty acids (popularly known as "trans fat"), a variety of fat that is rare in nature. Epidemiological evidence suggests that consumption of large amounts of saturated fatty acids and trans fat promotes obesity, type 2 diabetes, and atherosclerosis. The mechanism by which these fats exert these effects is under active investigation. Some evidence suggests that they promote an inflammatory response and may mute the action of insulin and other hormones (Section 27.3).

22.4 Fatty Acids Are Synthesized by Fatty Acid Synthase

Fatty acids are synthesized by a complex of enzymes that together are called *fatty acid synthase.* Because eating a typical Western diet meets our physiological needs for fats and lipids, adult human beings have little need for de novo fatty acid synthesis. However, many tissues, such as liver and adipose tissue, are capable of synthesizing fatty acids, and this synthesis is required under certain physiological conditions. For instance, fatty acid synthesis is necessary during embryonic development and during lactation in mammary glands. Inappropriate fatty acid synthesis in the liver of alcoholics contributes to liver failure.

Acetyl CoA, the end product of fatty acid degradation, is the precursor for virtually all fatty acids. The biochemical challenge is to link the two carbon units together and reduce the carbons to produce palmitate, a C_{16} fatty acid. Palmitate then serves as a precursor for the variety of other fatty acids.

Fatty acids are synthesized and degraded by different pathways

Although fatty acid synthesis is the reversal of the degradative pathway in regard to basic chemical reactions, the synthetic and degradative pathways are different mechanistically, again exemplifying the principle that *synthetic and degradative pathways are almost always distinct*. Some important differences between the pathways are as follows:

1. Synthesis takes place in the *cytoplasm,* in contrast with degradation, which takes place primarily in the mitochondrial matrix.

2. Intermediates in fatty acid synthesis are covalently linked to the sulfhydryl groups of an *acyl carrier protein* (ACP), whereas intermediates in fatty acid breakdown are covalently attached to the sulfhydryl group of coenzyme A.

3. The enzymes of fatty acid synthesis in higher organisms are joined in a *single polypeptide chain* called *fatty acid synthase.* In contrast, the degradative enzymes are not linked covalently.

4. The growing fatty acid chain is elongated by the *sequential addition of two-carbon units* derived from acetyl CoA. The activated donor of two-carbon units in the elongation step is *malonyl ACP.* The elongation reaction is driven by the release of CO_2.

5. The reductant in fatty acid synthesis is *NADPH,* whereas the oxidants in fatty acid degradation are NAD^+ and *FAD.*

6. The isomeric form of the hydroxyacyl intermediate in degradation is L, while the D form is used in synthesis.

The formation of malonyl CoA is the committed step in fatty acid synthesis

Fatty acid synthesis starts with the carboxylation of acetyl CoA to *malonyl CoA.* This irreversible reaction is the committed step in fatty acid synthesis.

The synthesis of malonyl CoA is catalyzed by the cytoplasmic enzyme *acetyl CoA carboxylase 1,* which contains a biotin prosthetic group. The carboxyl group of biotin is covalently attached to the ε amino group of a lysine residue, as in pyruvate carboxylase (Figure 16.24) and propionyl CoA carboxylase (p. 654). As with these other enzymes, a carboxybiotin intermediate is formed at the expense of the hydrolysis of a molecule of ATP. The activated CO_2 group in this intermediate is then transferred to acetyl CoA to form malonyl CoA.

$$\text{Biotin-enzyme} + \text{ATP} + \text{HCO}_3^- \rightleftharpoons \text{CO}_2\text{-biotin-enzyme} + \text{ADP} + \text{P}_i$$

$$\text{CO}_2\text{-biotin-enzyme} + \text{acetyl CoA} \longrightarrow \text{malonyl CoA} + \text{biotin-enzyme}$$

Acetyl CoA carboxylase 2, an isozyme of carboxylase 1 located in the mitochondria, is the essential regulatory enzyme for fatty acid metabolism (Section 22.5).

Intermediates in fatty acid synthesis are attached to an acyl carrier protein

The intermediates in fatty acid synthesis are linked to an acyl carrier protein. Specifically, they are linked to the sulfhydryl terminus of a phosphopantetheine group. In the degradation of fatty acids, this unit is present as part of coenzyme A, whereas, in their synthesis, it is attached to a serine residue of the acyl carrier protein (Figure 22.25). Thus, ACP, a single polypeptide chain of 77 residues, can be regarded as a giant prosthetic group, a "macro CoA."

Fatty acid synthesis consists of a series of condensation, reduction, dehydration, and reduction reactions

The enzyme system that catalyzes the synthesis of saturated long-chain fatty acids from acetyl CoA, malonyl CoA, and NADPH is called *fatty acid*

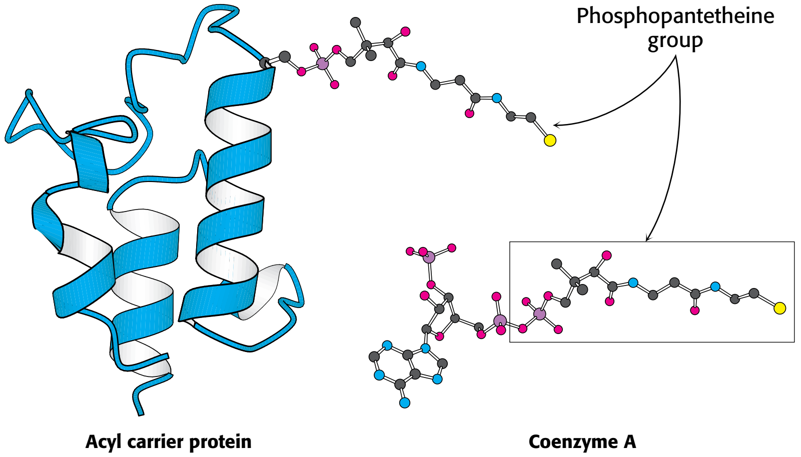


FIGURE 22.25 **Phosphopantetheine.** Both acyl carrier protein and coenzyme A include phosphopantetheine as their reactive units. [Drawn from 1ACP.pdb.]

TABLE 22.2 Principal reactions in fatty acid synthesis in bacteria

Step	Reaction	Enzyme
1	Acetyl CoA + HCO_3^- + ATP $\longrightarrow$ malonyl CoA + ADP + P_i + H^+	Acetyl CoA carboxylase
2	Acetyl CoA + ACP $\rightleftharpoons$ acetyl ACP + CoA	Acetyl transacylase
3	Malonyl CoA + ACP $\rightleftharpoons$ malonyl ACP + CoA	Malonyl transacylase
4	Acetyl ACP + malonyl ACP $\longrightarrow$ acetoacetyl ACP + ACP + CO_2	β-Ketoacyl synthase
5	Acetoacetyl ACP + NADPH + H^+ $\rightleftharpoons$ D-3-hydroxybutyryl ACP + $NADP^+$	β-Ketoacyl reductase
6	D-3-Hydroxybutyryl ACP $\rightleftharpoons$ crotonyl ACP + H_2O	3-Hydroxyacyl dehydratase
7	Crotonyl ACP + NADPH + H^+ $\longrightarrow$ butyryl ACP + $NADP^+$	Enoyl reductase

synthase. The synthase is actually a complex of distinct enzymes. The fatty acid synthase complex in bacteria is readily dissociated into individual enzymes when the cells are broken apart. The availability of these isolated enzymes has helped biochemists elucidate the steps in fatty acid synthesis (Table 22.2). In fact, the reactions leading to fatty acid synthesis in higher organisms are very much like those of bacteria.

The elongation phase of fatty acid synthesis starts with the formation of acetyl ACP and malonyl ACP. *Acetyl transacylase* and *malonyl transacylase* catalyze these reactions.

$$\text{Acetyl CoA} + \text{ACP} \rightleftharpoons \text{acetyl ACP} + \text{CoA}$$
$$\text{Malonyl CoA} + \text{ACP} \rightleftharpoons \text{malonyl ACP} + \text{CoA}$$

Malonyl transacylase is highly specific, whereas acetyl transacylase can transfer acyl groups other than the acetyl unit, though at a much slower rate. The synthesis of fatty acids with an odd number of carbon atoms starts with propionyl ACP, which is formed from propionyl CoA by acetyl transacylase.

Acetyl ACP and malonyl ACP react to form acetoacetyl ACP (Figure 22.26). The *β-ketoacyl synthase,* also called the condensing enzyme, catalyzes this condensation reaction.

$$\text{Acetyl ACP} + \text{malonyl ACP} \longrightarrow \text{acetoacetyl ACP} + \text{ACP} + CO_2$$

In the condensation reaction, a four-carbon unit is formed from a two-carbon unit and a three-carbon unit, and CO_2 is released. Why is the four-carbon unit not formed from two 2-carbon units—say, two molecules of acetyl ACP? The answer is that the equilibrium for the synthesis of acetoacetyl ACP from two molecules of acetyl ACP is highly unfavorable. In contrast, *the equilibrium is favorable if malonyl ACP is a reactant because its decarboxylation contributes a substantial decrease in free energy.* In effect, ATP drives the condensation reaction, though ATP does not directly participate in the condensation reaction. Instead, ATP is used to carboxylate acetyl CoA to malonyl CoA. The free energy thus stored in malonyl CoA is released in the decarboxylation accompanying the formation of acetoacetyl ACP. Although HCO_3^- is required for fatty acid synthesis, its carbon atom does not appear in the product. Rather, *all the carbon atoms of fatty acids containing an even number of carbon atoms are derived from acetyl CoA.*

The next three steps in fatty acid synthesis reduce the keto group at C-3 to a methylene group (Figure 22.26). First, acetoacetyl ACP is reduced to D-3-hydroxybutyryl ACP by β-ketoacyl reductase. This reaction differs from the corresponding one in fatty acid degradation in two respects: (1) the D rather than the L isomer is formed; and (2) NADPH is the reducing agent, whereas NAD^+ is the oxidizing agent in β oxidation. This difference exemplifies the general principle that *NADPH is consumed in biosynthetic reactions, whereas NADH is generated in energy-yielding reactions.* Then, D-3-hydroxybutyryl ACP is *dehydrated* to form crotonyl ACP, which is a

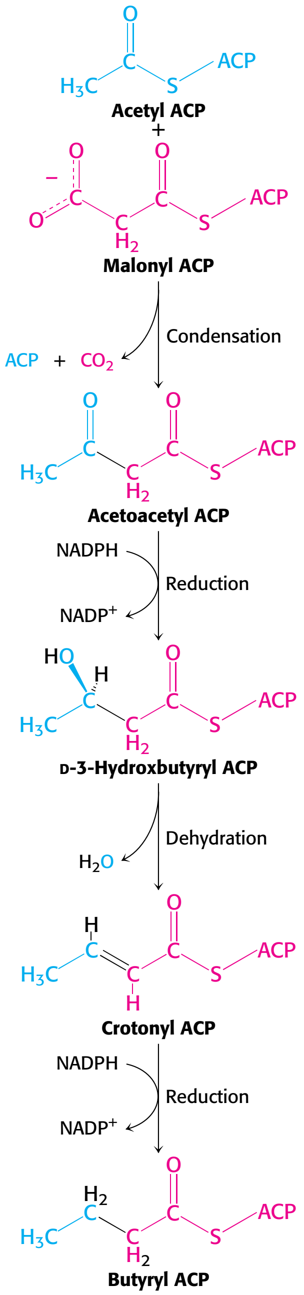


FIGURE 22.26 The steps of fatty acid synthesis. Fatty acid synthesis begins with the condensation of malonyl ACP and acetyl ACP to form acetoacetyl ACP. Acetoacetyl ACP is then reduced, dehydrated, and reduced again to form butyryl ACP. Another cycle begins with the condensation of butyryl ACP and malonyl ACP. The sequence of reactions is repeated until the final product palmitate is formed.

trans-Δ^2-enoyl ACP by 3-hydroxyacyl dehydratase. The final step in the cycle *reduces* crotonyl ACP to butyryl ACP. NADPH is again the reductant, whereas FAD is the oxidant in the corresponding reaction in β oxidation. The bacterial enzyme that catalyzes this step, *enoyl reductase,* can be inhibited by *triclosan,* a broad-spectrum antibacterial agent that is added to a variety of products such as toothpaste, soaps, and skin creams. These last three reactions—a reduction, a dehydration, and a second reduction—convert acetoacetyl ACP into butyryl ACP, which completes the first elongation cycle.

In the second round of fatty acid synthesis, butyryl ACP condenses with malonyl ACP to form a C_6-β-ketoacyl ACP. This reaction is like the one in the first round, in which acetyl ACP condenses with malonyl ACP to form a C_4-β-ketoacyl ACP. Reduction, dehydration, and a second reduction convert the C_6-β-ketoacyl ACP into a C_6-acyl ACP, which is ready for a third round of elongation. The elongation cycles continue until C_{16}-acyl ACP is formed. This intermediate is a good substrate for a thioesterase that hydrolyzes C_{16}-acyl ACP to yield palmitate and ACP. *The thioesterase acts as a ruler to determine fatty acid chain length.* The synthesis of longer-chain fatty acids is discussed in Section 22.6.

Fatty acids are synthesized by a multifunctional enzyme complex in animals

Although the basic biochemical reactions in fatty acid synthesis are very similar in *E. coli* and eukaryotes, the structure of the synthase varies considerably. The component enzymes of animal fatty acid synthases, in contrast with those of *E. coli* and plants, are linked in a large polypeptide chain.

The structure of a large part of the mammalian fatty acid synthase has recently been determined, with the acyl carrier protein and thioesterase remaining to be resolved. The enzyme is a dimer of identical 270-kd subunits. Each chain contains all of the active sites required for activity, as well as an acyl carrier protein tethered to the complex (Figure 22.27A). Despite the fact that each chain possesses all of the enzymes required for fatty acid synthesis, the monomers are not active. A dimer is required.

The two component chains interact such that the enzyme activities are partitioned into two distinct compartments (Figure 22.27B). The selecting and condensing compartment binds the acetyl and malonyl substrates and condenses them to form the growing chain. Interestingly, the mammalian fatty acid synthase has one active site, malonyl/acetyl transacylase, that adds both acetyl CoA and malonyl CoA. In contrast, most other fatty acid synthases have two separate enzyme activities, one for acetyl CoA and one for malonyl CoA. The modification compartment is responsible for the reduction and dehydration activities that result in the saturated fatty acid product.

Let us consider one catalytic cycle of the fatty acid synthase complex (Figure 22.28). An elongation cycle begins when malonyl/acetyl transacylase (MAT) moves an acetyl unit from coenzyme A to the acyl carrier protein (ACP). β-Ketosynthase (β-KS) accepts the acetyl unit, which forms a thioester with a cysteine residue at the β-KS active site. The vacant ACP is reloaded by MAT, this time with a malonyl moiety. Malonyl ACP visits the active site of β-KS where the condensation of the two 2-carbon fragments takes place on the ACP with the concomitant release of CO_2. The selecting and condensing process concludes with the β-ketoacyl product attached to the ACP.

The loaded ACP then sequentially visits the active sites of the modification compartment of the enzyme, where the β-keto group of the substrate is reduced to —OH, dehydrated, and finally reduced to yield the saturated acyl product, still attached to the ACP. With the completion of the modification

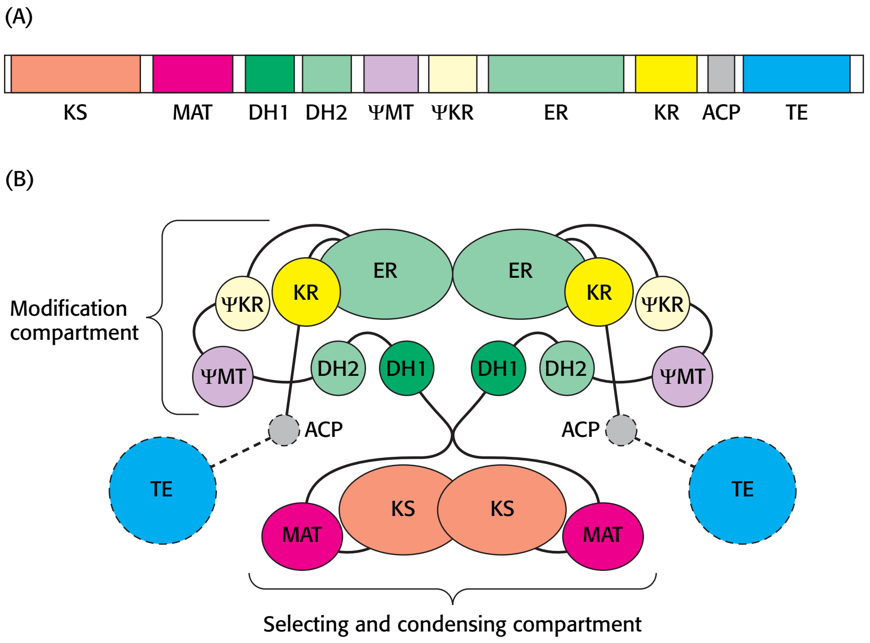

(A)
KS
MAT
DH1
DH2
ΨMT
ΨKR
ER
KR
ACP
TE
(B)
Modification compartment
ER
ER
KR
KR
ΨKR
ΨKR
ΨMT
ΨMT
DH2
DH1
DH1
DH2
ACP
ACP
TE
TE
KS
KS
MAT
MAT
Selecting and condensing compartment

FIGURE 22.27 The structure of the mammalian fatty acid synthase. (A) The arrangement of the catalytic activities present in a single polypeptide chain. (B) A cartoon of the dimer based on an x-ray crystallographic result. The Ψ-MT and Ψ-KR are inactive domains similar to methyl transferase and ketoreductase sequences. Although there are two domains for DH, only one is active. The inactive domains are presented in faded colors. Dotted lines outline domains for which the structure has not yet been determined. Abbreviations: KS, ketosynthase; MAT, malonylacetyl transferase; DH, dehydratase; Ψ-MT, methyl transferase (inactive); Ψ-KR, ketoreductase (inactive); ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; TE, thioesterase.

process, the reduced product is transferred to the β-KS while the ACP accepts another malonyl unit. Condensation takes place and is followed by another modification cycle. The process is repeated until the thioesterase releases the final C_{16} palmitic acid product.

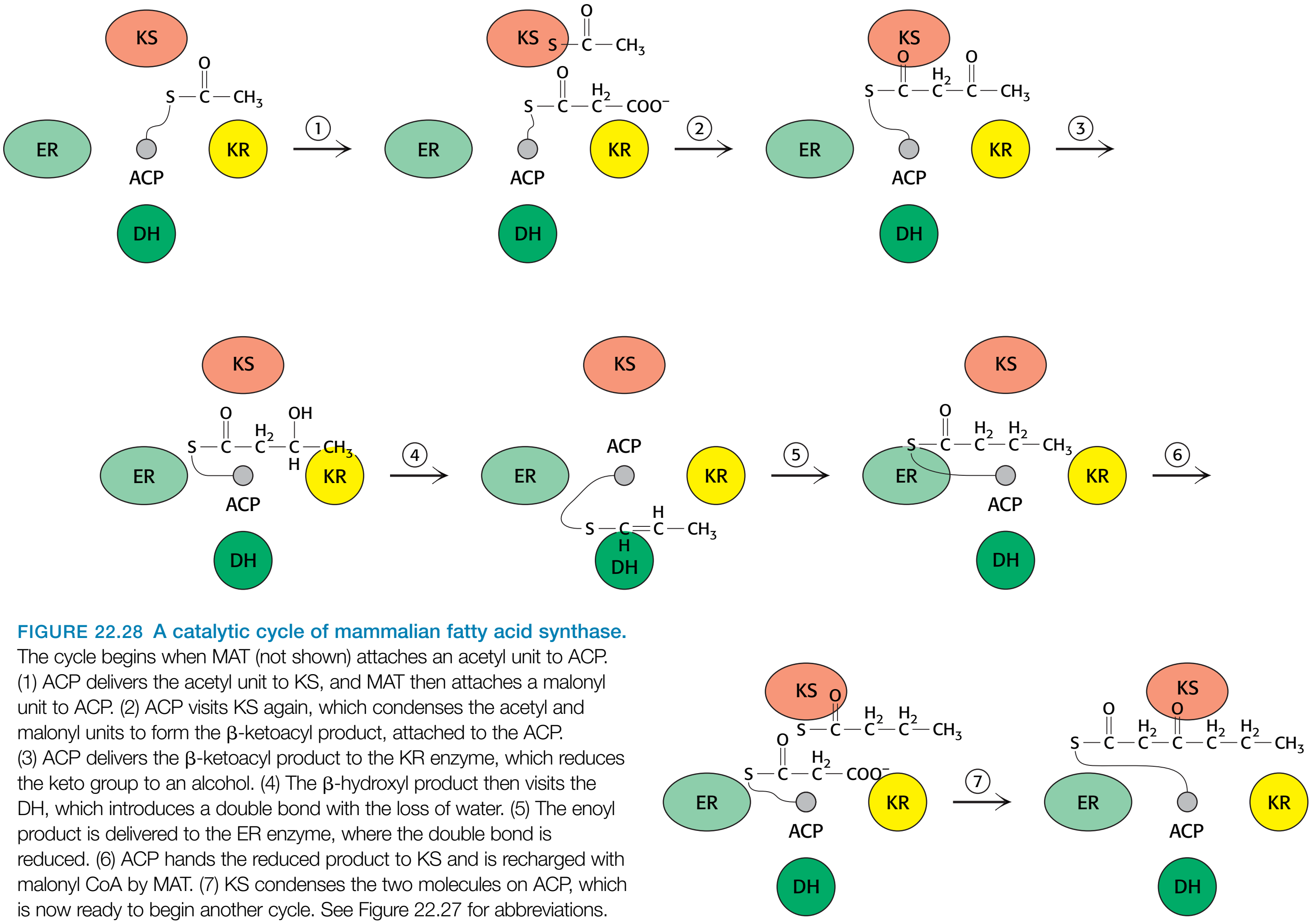

KS
ER
KR
ACP
DH
S—C(=O)—CH3
①
②
③
④
⑤
⑥
⑦
S—C(=O)—CH2—COO⁻
S—C(=O)—CH2—C(=O)—CH3
S—C(=O)—CH2—CH(OH)—CH3
S—CH=CH—CH3
S—C(=O)—CH2—CH2—CH3
S—C(=O)—CH2—C(=O)—CH2—CH2—CH3

FIGURE 22.28 A catalytic cycle of mammalian fatty acid synthase. The cycle begins when MAT (not shown) attaches an acetyl unit to ACP. (1) ACP delivers the acetyl unit to KS, and MAT then attaches a malonyl unit to ACP. (2) ACP visits KS again, which condenses the acetyl and malonyl units to form the β-ketoacyl product, attached to the ACP. (3) ACP delivers the β-ketoacyl product to the KR enzyme, which reduces the keto group to an alcohol. (4) The β-hydroxyl product then visits the DH, which introduces a double bond with the loss of water. (5) The enoyl product is delivered to the ER enzyme, where the double bond is reduced. (6) ACP hands the reduced product to KS and is recharged with malonyl CoA by MAT. (7) KS condenses the two molecules on ACP, which is now ready to begin another cycle. See Figure 22.27 for abbreviations.

Many eukaryotic multienzyme complexes are multifunctional proteins in which different enzymes are linked covalently. Multifunctional enzymes such as fatty acid synthase seem likely to have arisen in eukaryotic evolution by fusion of the individual genes of evolutionary ancestors.

The synthesis of palmitate requires 8 molecules of acetyl CoA, 14 molecules of NADPH, and 7 molecules of ATP

The stoichiometry of the synthesis of palmitate is

$$\text{Acetyl CoA} + 7\text{ malonyl CoA} + 14\text{ NADPH} + 20\text{ H}^+ \longrightarrow \text{palmitate} + 7\text{ CO}_2 + 14\text{ NADP}^+ + 8\text{ CoA} + 6\text{ H}_2\text{O}$$

The equation for the synthesis of the malonyl CoA used in the preceding reaction is

$$7\text{ Acetyl CoA} + 7\text{ CO}_2 + 7\text{ ATP} \longrightarrow 7\text{ malonyl CoA} + 7\text{ ADP} + 7\text{ P}_i + 14\text{ H}^+$$

Hence, the overall stoichiometry for the synthesis of palmitate is

$$8\text{ Acetyl CoA} + 7\text{ ATP} + 14\text{ NADPH} + 6\text{ H}^+ \longrightarrow \text{palmitate} + 14\text{ NADP}^+ + 8\text{ CoA} + 6\text{ H}_2\text{O} + 7\text{ ADP} + 7\text{ P}_i$$

Citrate carries acetyl groups from mitochondria to the cytoplasm for fatty acid synthesis

Fatty acids are synthesized in the cytoplasm, whereas acetyl CoA is formed from pyruvate in mitochondria. Hence, acetyl CoA must be transferred from mitochondria to the cytoplasm for fatty acid synthesis. Mitochondria, however, are not readily permeable to acetyl CoA. Recall that carnitine carries only long-chain fatty acids. *The barrier to acetyl CoA is bypassed by citrate, which carries acetyl groups across the inner mitochondrial membrane.* Citrate is formed in the mitochondrial matrix by the condensation of acetyl CoA with oxaloacetate (Figure 22.29). When present at high levels, citrate is transported to the cytoplasm, where it is cleaved by *ATP-citrate lyase.*

$$\text{Citrate} + \text{ATP} + \text{CoA} + \text{H}_2\text{O} \longrightarrow \text{acetyl CoA} + \text{ADP} + \text{P}_i + \text{oxaloacetate}$$

Lyases
Enzymes catalyzing the cleavage of C—C, C—O, or C—N bonds by elimination. A double bond is formed in these reactions.

This reaction occurs in three steps: (1) The formation of a phospho-enzyme with the donation of a phosphoryl group from ATP; (2) binding of citrate and CoA followed by the formation of citroyl CoA and release of the phosphate; (3) cleavage of citroyl CoA to yield acetyl CoA and oxaloacetate. As we will see shortly (Section 22.6), citrate stimulates acetyl CoA carboxylase, the enzyme that regulates fatty acid metabolism. Recall also that the

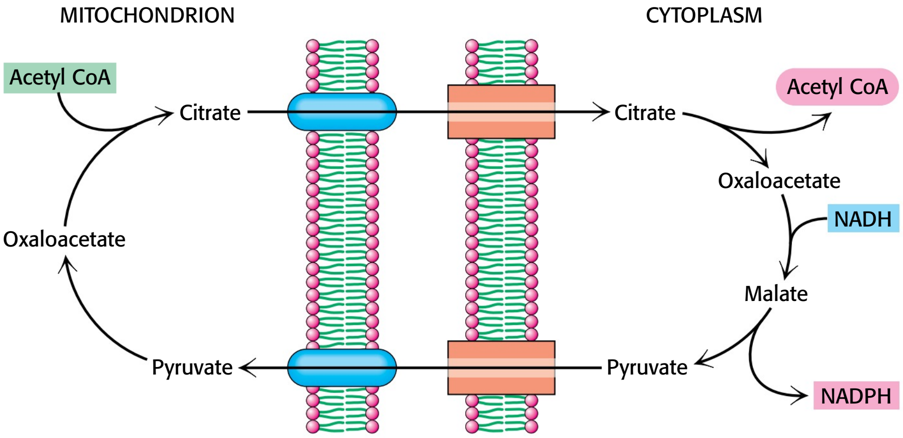


FIGURE 22.29 Transfer of acetyl CoA to the cytoplasm. Acetyl CoA is transferred from mitochondria to the cytoplasm, and the reducing potential of NADH is concomitantly converted into that of NADPH by this series of reactions.

presence of citrate in the cytoplasm inhibits phosphofructokinase, the enzyme that controls the glycolytic pathway.

ATP-citrate lyase is stimulated by insulin, which initiates a signal transduction pathway that ultimately results in the phosphorylation and activation of the lyase by protein kinase B (also called Akt).

Several sources supply NADPH for fatty acid synthesis

Oxaloacetate formed in the transfer of acetyl groups to the cytoplasm must now be returned to the mitochondria. The inner mitochondrial membrane is impermeable to oxaloacetate. Hence, a series of bypass reactions are needed. These reactions also generate much of the NADPH needed for fatty acid synthesis. First, oxaloacetate is reduced to malate by NADH. This reaction is catalyzed by a *malate dehydrogenase* in the cytoplasm.

$$\text{Oxaloacetate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{malate} + \text{NAD}^+$$

Second, malate is oxidatively decarboxylated by an *NADP⁺-linked malate enzyme* (also called *malic enzyme*).

$$\text{Malate} + \text{NADP}^+ \longrightarrow \text{pyruvate} + \text{CO}_2 + \text{NADPH}$$

The pyruvate formed in this reaction readily enters mitochondria, where it is carboxylated to oxaloacetate by pyruvate carboxylase.

$$\text{Pyruvate} + \text{CO}_2 + \text{ATP} + \text{H}_2\text{O} \longrightarrow \text{oxaloacetate} + \text{ADP} + \text{P}_\text{i} + 2\,\text{H}^+$$

The sum of these three reactions is

$$\text{NADP}^+ + \text{NADH} + \text{ATP} + \text{H}_2\text{O} \longrightarrow \text{NADPH} + \text{NAD}^+ + \text{ADP} + \text{P}_\text{i} + \text{H}^+$$

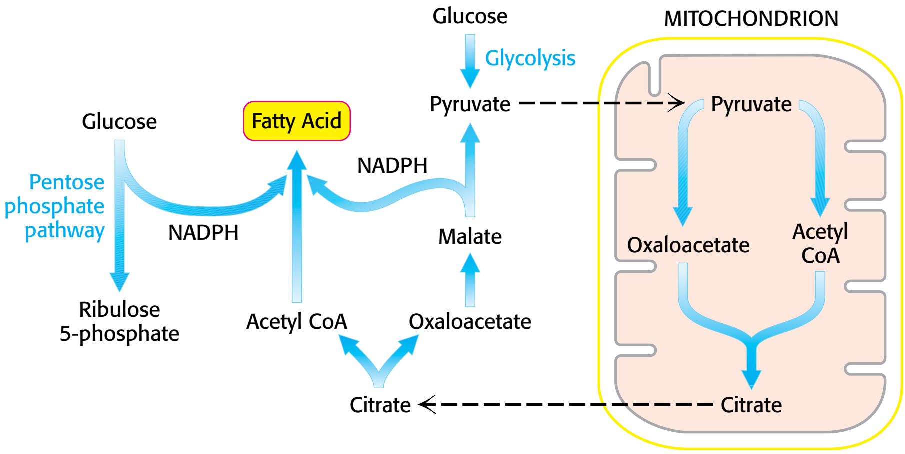
Thus, *one molecule of NADPH is generated for each molecule of acetyl CoA that is transferred from mitochondria to the cytoplasm.* Hence, eight molecules of NADPH are formed when eight molecules of acetyl CoA are transferred to the cytoplasm for the synthesis of palmitate. *The additional six molecules of NADPH required for this process come from the pentose phosphate pathway* (Section 20.3).

The accumulation of the precursors for fatty acid synthesis is a wonderful example of the coordinated use of multiple pathways. The citric acid cycle, transport of oxaloacetate from the mitochondria, and pentose phosphate pathway provide the carbon atoms and reducing power, whereas glycolysis and oxidative phosphorylation provide the ATP to meet the needs for fatty acid synthesis (Figure 22.30).

Fatty acid metabolism is altered in tumor cells

We have previously seen that cancer cells alter glucose metabolism to meet the needs of rapid cell growth. Cancer cells must also increase fatty acid synthesis for use as signal molecules as well as for incorporation into membrane phospholipids. Many of the enzymes of fatty acid synthesis are overexpressed in most human cancers, and this expression is correlated with tumor malignancy. Recall that normal cells do little de novo fatty acid synthesis, relying instead on dietary intake to meet their fatty needs.

FIGURE 22.30 PATHWAY INTEGRATION: Fatty acid synthesis. Fatty acid synthesis requires the cooperation of various metabolic pathways located in different cellular compartments.



The dependence of de novo fatty acid synthesis provides possible therapeutic targets to inhibit cancer cell growth. Inhibition of β-ketoacyl ACP synthase, the enzyme that catalyzes the condensation step of fatty acid synthesis, does indeed inhibit phospholipid synthesis and subsequent cell growth in some cancers, apparently by inducing apoptosis. However, another startling observation was made: *mice treated with inhibitors of the β-ketoacyl ACP synthase showed remarkable weight loss* because they ate less. Thus, fatty acid synthase inhibitors are exciting candidates both as antitumor and as antiobesity drugs.

Acetyl CoA carboxylase is also being investigated as a possible target for inhibiting cancer cell growth. Inhibition of the carboxylase in prostate and breast cancer cell lines induces apoptosis in the cancer cells, and yet is without effect in normal cells (problem 54). Understanding the alteration of fatty acid metabolism in cancer cells is a developing area of research that holds promise of generating new cancer therapies.

22.5 The Elongation and Unsaturation of Fatty Acids Are Accomplished by Accessory Enzyme Systems

The major product of the fatty acid synthase is palmitate. In eukaryotes, longer fatty acids are formed by elongation reactions catalyzed by enzymes on the cytoplasmic face of the *endoplasmic reticulum membrane*. These reactions add two-carbon units sequentially to the carboxyl ends of both saturated and unsaturated fatty acyl CoA substrates. Malonyl CoA is the two-carbon donor in the elongation of fatty acyl CoAs. Again, condensation is driven by the decarboxylation of malonyl CoA.

Membrane-bound enzymes generate unsaturated fatty acids

Endoplasmic reticulum systems also introduce double bonds into long-chain acyl CoAs. For example, in the conversion of stearoyl CoA into oleoyl CoA, a *cis*-Δ^9 double bond is inserted by an oxidase that employs *molecular oxygen* and *NADH* (or *NADPH*).

$$\text{Stearoyl CoA} + \text{NADH} + \text{H}^+ + \text{O}_2 \longrightarrow \text{oleoyl CoA} + \text{NAD}^+ + 2\,\text{H}_2\text{O}$$

This reaction is catalyzed by a complex of three membrane-bound proteins: *NADH-cytochrome* b_5 *reductase, cytochrome* b_5, *and stearoyl CoA desaturase* (Figure 22.31). First, electrons are transferred from NADH to the FAD moiety of NADH-cytochrome b_5 reductase. The heme iron atom of cytochrome b_5 is then reduced to the Fe^{2+} state. The nonheme iron atom of the desaturase is subsequently converted into the Fe^{2+} state, which enables it to interact with O_2 and the saturated fatty acyl CoA substrate. A double bond is formed and two molecules of H_2O are released. Two electrons come from NADH and two from the single bond of the fatty acyl substrate.

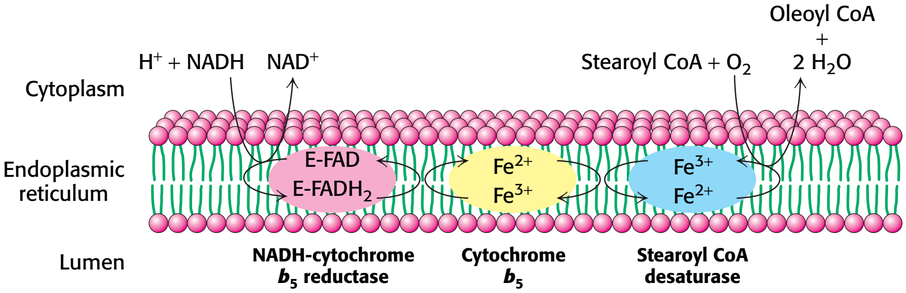


FIGURE 22.31 Electron-transport chain in the desaturation of fatty acids.

A variety of unsaturated fatty acids can be formed from oleate by a combination of elongation and desaturation reactions. For example, oleate can be elongated to a 20:1 *cis*-Δ^{11} fatty acid. Alternatively, a second double bond can be inserted to yield an 18:2 *cis*-Δ^6, Δ^9 fatty acid. Similarly, palmitate (16:0) can be oxidized to palmitoleate (16:1 *cis*-Δ^9), which can then be elongated to *cis*-vaccenate (18:1 *cis*-Δ^{11}).

Unsaturated fatty acids in mammals are derived from either palmitoleate (16:1), oleate (18:1), linoleate (18:2), or linolenate (18:3). The number of carbon atoms from the ω end of a derived unsaturated fatty acid to the nearest double bond identifies its precursor.

Precursor	Formula
Linolenate (ω-3)	$CH_3{-}(CH_2)_2{=}CH{-}R$
Linoleate (ω-6)	$CH_3{-}(CH_2)_5{=}CH{-}R$
Palmitoleate (ω-7)	$CH_3{-}(CH_2)_6{=}CH{-}R$
Oleate (ω-9)	$CH_3{-}(CH_2)_8{=}CH{-}R$

Mammals lack the enzymes to introduce double bonds at carbon atoms beyond C-9 in the fatty acid chain. Hence, mammals cannot synthesize linoleate (18:2 *cis*-Δ^9, Δ^{12}) and linolenate (18:3 *cis*-Δ^9, Δ^{12}, Δ^{15}). Linoleate and linolenate are the two essential fatty acids. The term *essential* means that they must be supplied in the diet because they are required by an organism and cannot be synthesized by the organism itself. Linoleate and linolenate furnished by the diet are the starting points for the synthesis of a variety of other unsaturated fatty acids.

Eicosanoid hormones are derived from polyunsaturated fatty acids

Arachidonate, a 20:4 fatty acid derived from linoleate, is the major precursor of several classes of signal molecules: prostaglandins, prostacyclins, thromboxanes, and leukotrienes (Figure 22.32).

A prostaglandin is a 20-carbon fatty acid containing a 5-carbon ring (Figure 22.33). This basic compound is modified by reductases and isomerases to yield nine major classes of prostaglandins, designated PGA through PGI; a subscript denotes the number of carbon–carbon double bonds outside the ring. Prostaglandins with two double bonds, such as PGE_2, are derived from arachidonate; the other two double bonds of this precursor are lost in forming a 5-membered ring. *Prostacyclin* and *thromboxanes* are related compounds that arise from a nascent prostaglandin. They are generated by *prostacyclin synthase* and *thromboxane synthase,* respectively. Alternatively, arachidonate can be converted into *leukotrienes* by the action of *lipoxygenase.* Leukotrienes, first found in leukocytes, contain three conjugated double bonds—hence the name. Prostaglandins, prostacyclin, thromboxanes, and leukotrienes are called *eicosanoids* (from the Greek *eikosi,* "twenty") because they contain 20 carbon atoms.

Prostaglandins and other eicosanoids are *local hormones* because they are short-lived. They alter the activities both of the cells in which they are synthesized and of adjoining cells by binding to 7TM receptors. Their effects may vary from one cell type to another, in contrast with the more-uniform actions of global hormones such as insulin and glucagon. Prostaglandins stimulate inflammation, regulate blood flow to particular organs, control ion transport across membranes, modulate synaptic transmission, and induce sleep.

Recall that aspirin blocks access to the active site of the enzyme that converts arachidonate into prostaglandin H_2 (Section 12.3). Because arachidonate is the precursor of other prostaglandins, prostacyclin, and thromboxanes, blocking this step interferes with many signaling pathways. Aspirin's ability to obstruct these pathways accounts for its wide-ranging effects on inflammation, fever, pain, and blood clotting.

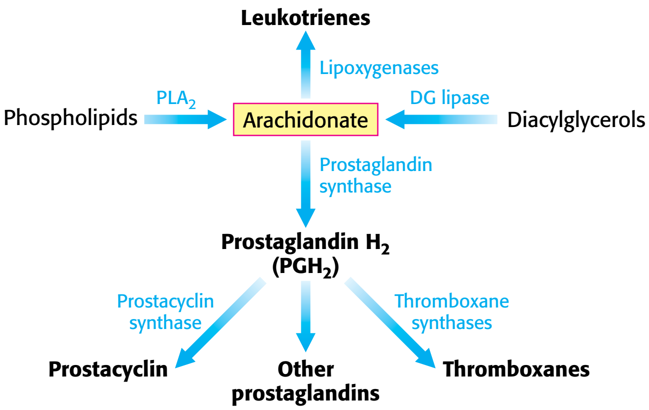
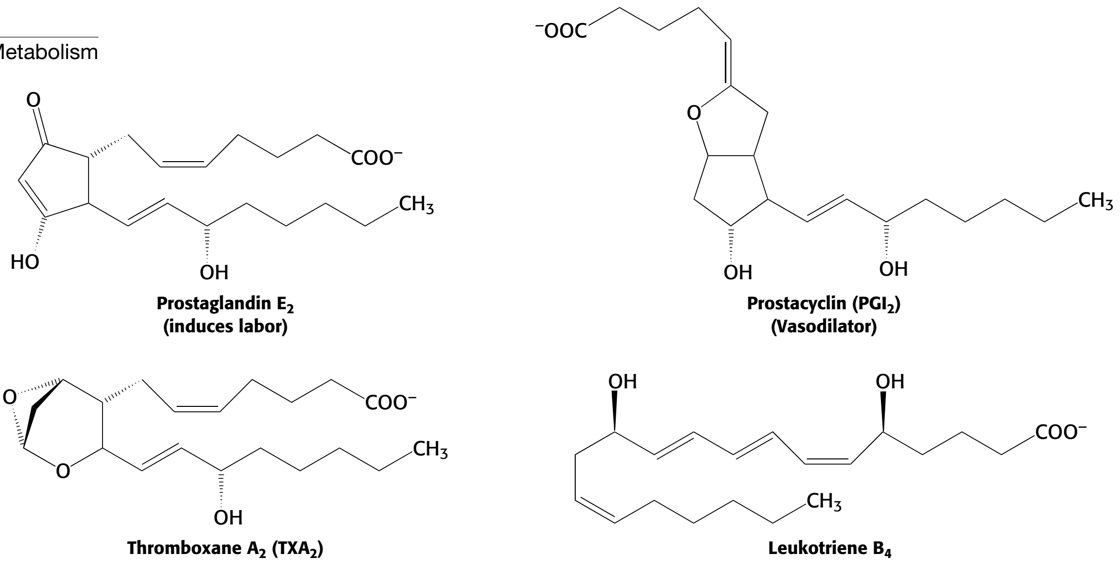


FIGURE 22.32 Arachidonate is the major precursor of eicosanoid hormones. Prostaglandin synthase catalyzes the first step in a pathway leading to prostaglandins, prostacyclins, and thromboxanes. Lipoxygenase catalyzes the initial step in a pathway leading to leukotrienes.



Prostaglandin E_2
(induces labor)

Prostacyclin (PGI_2)
(Vasodilator)

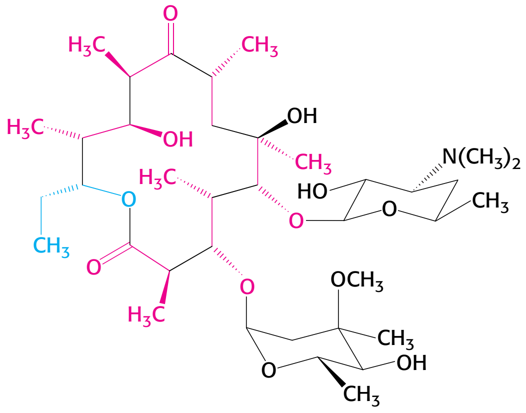
Thromboxane A_2 (TXA_2)
(Enhances platelet aggregation)

Leukotriene B_4
(Proinflammatory signal)

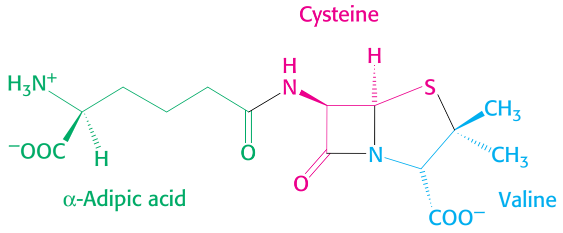
FIGURE 22.33 Structures of several eicosanoids.

Variations on a theme: Polyketide and nonribosomal peptide synthetases resemble fatty acid synthase

The mammalian multifunctional fatty acid synthase is a member of a large family of complex enzymes termed *megasynthases* that participate in step-by-step synthetic pathways. Two important classes of compounds that are synthesized by such enzymes are the *polyketides* and the *nonribosomal peptides*. These classes of compounds provide a variety of useful drugs, including antibiotics, immunosuppressants, antifungal agents and anticancer drugs. The antibiotic erythromycin is an example of a polyketide, whereas penicillin (Section 8.5) is a nonribosomal peptide.



Erythromycin



Penicillin

Research is underway to learn how to manipulate the pathways and the enzymes of polyketide and nonribosomal peptide synthesis in order to generate new therapeutics.

22.6 Acetyl CoA Carboxylase Plays a Key Role in Controlling Fatty Acid Metabolism

Fatty acid metabolism is stringently controlled so that synthesis and degradation are highly responsive to physiological needs. Fatty acid synthesis is maximal when carbohydrates and energy are plentiful and when fatty acids are scarce. *Acetyl CoA carboxylase 1 and 2 play essential roles in regulating fatty acid synthesis and degradation.* Recall that this enzyme catalyzes the committed step in fatty acid synthesis: the production of malonyl CoA (the activated two-carbon donor). This important enzyme is subject to both local and hormonal regulation. We will examine each of these levels of regulation in turn.

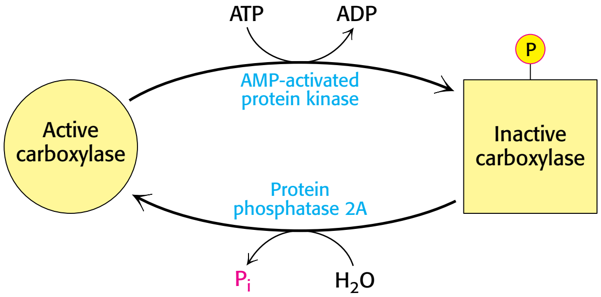


FIGURE 22.34 Control of acetyl CoA carboxylase. Acetyl CoA carboxylase is inhibited by phosphorylation.

Acetyl CoA carboxylase is regulated by conditions in the cell

Acetyl CoA carboxylase responds to changes in its immediate environment. *Acetyl CoA carboxylase is switched off by phosphorylation* and activated by dephosphorylation (Figure 22.34). *AMP-activated protein kinase (AMPK)* converts the carboxylase into an inactive form by modifying three serine residues. AMPK is essentially a fuel gauge; it is activated by AMP and inhibited by ATP.

The carboxylase is also allosterically stimulated by citrate. The level of citrate is high when both acetyl CoA and ATP are abundant, signifying that raw materials and energy are available for fatty acid synthesis. Citrate acts in an unusual manner on inactive acetyl CoA carboxylase, which exists as isolated inactive dimers. Citrate facilitates the polymerization of the inactive dimers into active filaments (Figure 22.35). However, polymerization by citrate alone requires supraphysiological concentrations. In the cell, citrate-induced polymerization is facilitated by the protein MIG12, which greatly reduces the amount of citrate required. Polymerization can partly reverse the inhibition produced by phosphorylation (Figure 22.36). The stimulatory effect of citrate on the carboxylase is counteracted by *palmitoyl CoA,* which is abundant when there is an excess of fatty acids. Palmitoyl CoA causes the filaments to disassemble into the inactive subunits. Palmitoyl CoA also inhibits the translocase that transports citrate from mitochondria to the cytoplasm, as well as glucose 6-phosphate dehydrogenase, which generates NADPH in the pentose phosphate pathway.

The isozyme acetyl CoA carboxylase 2, located in the mitochondria, plays a role in the regulation of fatty acid degradation. Malonyl CoA, the product of the carboxylase reaction, is present at a high level when fuel molecules are abundant. *Malonyl CoA inhibits carnitine acyltransferase I, preventing the entry of fatty acyl CoAs into the mitochondrial matrix in times of plenty.* Malonyl CoA is an especially effective inhibitor of carnitine acyltransferase I in heart and muscle, tissues that have little fatty acid synthesis capacity of their own. In these tissues, acetyl CoA carboxylase may be a purely regulatory enzyme.

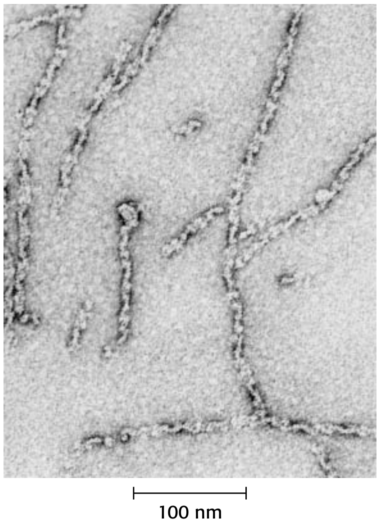


FIGURE 22.35 Filaments of acetyl CoA carboxylase. The electron micrograph shows the enzymatically active filamentous form of acetyl CoA carboxylase from chicken liver. The inactive form is a dimer of 265-kDa subunits. [Courtesy of Dr. M. Daniel Lane.]

Acetyl CoA carboxylase is regulated by a variety of hormones

Acetyl CoA carboxylase is controlled by the hormones glucagon, epinephrine, and insulin, which denote the overall energy status of the organism. *Insulin stimulates fatty acid synthesis by activating the carboxylase, whereas glucagon and epinephrine have the reverse effect.*

Regulation by glucagon and epinephrine. Consider, again, a person who has just awakened from a night's sleep and begins a bout of exercise. As mentioned, glycogen stores will be low, but lipids are readily available for mobilization.

As stated earlier, the hormones glucagon and epinephrine, present under conditions of fasting and exercise, will stimulate the release of fatty acids from triacylglycerols in fat cells, which will be released into the blood, and probably from muscle cells, where they will be used immediately as fuel. These same hormones will inhibit fatty acid synthesis by inhibiting

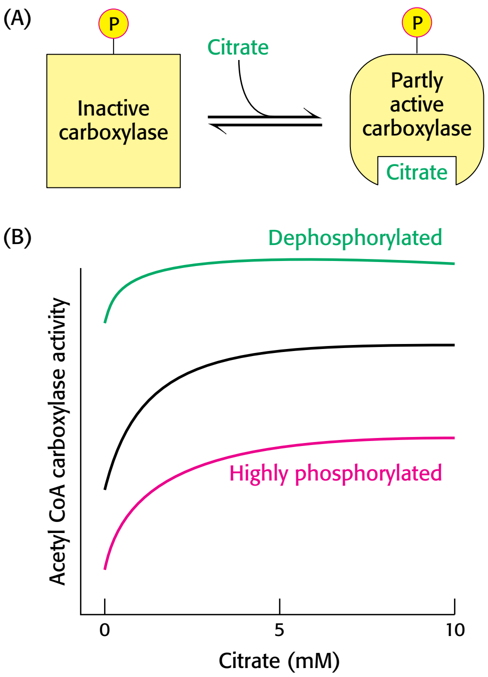


FIGURE 22.36 Dependence of the catalytic activity of acetyl CoA carboxylase on the concentration of citrate. (A) Citrate can partly activate the phosphorylated carboxylase. (B) The dephosphorylated form of the carboxylase is highly active even when citrate is absent. Citrate partly overcomes the inhibition produced by phosphorylation. [Information from G. M. Mabrouk, I. M. Helmy, K. G. Thampy, and S. J. Wakil. *J. Biol. Chem.* 265:6330–6338, 1990.]

acetyl CoA carboxylase. Although the exact mechanism by which these hormones exert their effects is not known, the net result is to augment the inhibition by the AMP-activated kinase. This result makes sound physiological sense: when the energy level of the cell is low, as signified by a high concentration of AMP, and the energy level of the organism is low, as signaled by glucagon, fats should not be synthesized. Epinephrine, which signals the need for immediate energy, enhances this effect. Hence, *these catabolic hormones switch off fatty acid synthesis by keeping the carboxylase in the inactive phosphorylated state.*

Regulation by insulin. Now consider the situation after the exercise has ended and the runner has had a meal. In this case, the hormone insulin inhibits the mobilization of fatty acids and stimulates their accumulation as triacylglycerols by muscle and adipose tissue. Insulin also stimulates fatty acid synthesis by activating acetyl CoA carboxylase. Insulin activates the carboxylase by enhancing the phosphorylation of AMPK by protein kinase B, which inhibits AMPK, as well as by stimulating the activity of a protein phosphatase that dephosphorylates and activates acetyl CoA carboxylase. Thus, the signal molecules glucagon, epinephrine, and insulin act in concert on triacylglycerol metabolism and acetyl CoA carboxylase to carefully regulate the utilization and storage of fatty acids.

Response to diet. *Long-term control is mediated by changes in the rates of synthesis and degradation of the enzymes participating in fatty acid synthesis.* Animals that have fasted and are then fed high-carbohydrate, low-fat diets show marked increases in their amounts of acetyl CoA carboxylase and fatty acid synthase within a few days. This type of regulation is known as *adaptive control*. This regulation, which is mediated both by insulin and by glucose, is at the level of gene transcription.

SUMMARY

22.1 Triacylglycerols Are Highly Concentrated Energy Stores

Fatty acids are physiologically important as (1) fuel molecules, (2) components of phospholipids and glycolipids, (3) hydrophobic modifiers of proteins, and (4) hormones and intracellular messengers. They are stored in adipose tissue as triacylglycerols (neutral fat).

22.2 The Use of Fatty Acids as Fuel Requires Three Stages of Processing

Triacylglycerols can be mobilized by the hydrolytic action of lipases that are under hormonal control. Glucagon and epinephrine stimulate triacylglycerol breakdown by activating the lipase. Insulin, in contrast, inhibits lipolysis. Fatty acids are activated to acyl CoAs, transported across the inner mitochondrial membrane by carnitine, and degraded in the mitochondrial matrix by a recurring sequence of four reactions: oxidation by FAD, hydration, oxidation by NAD^+, and thiolysis by coenzyme A. The $FADH_2$ and NADH formed in the oxidation steps transfer their electrons to O_2 by means of the respiratory chain, whereas the acetyl CoA formed in the thiolysis step normally enters the citric acid cycle by condensing with oxaloacetate.

22.3 Unsaturated and Odd-Chain Fatty Acids Require Additional Steps for Degradation

Fatty acids that contain double bonds or odd numbers of carbon atoms require ancillary steps to be degraded. An isomerase and a reductase are required for the oxidation of unsaturated fatty acids, whereas propionyl

CoA derived from chains with odd numbers of carbon atoms requires a vitamin B_{12}-dependent enzyme to be converted into succinyl CoA. Ketone bodies, formed from acetyl CoA, are an important fuel source for some tissues. Mammals are unable to convert fatty acids into glucose, because they lack a pathway for the net production of oxaloacetate, pyruvate, or other gluconeogenic intermediates from acetyl CoA.

22.4 Fatty Acids Are Synthesized by Fatty Acid Synthase

Fatty acids are synthesized in the cytoplasm by a different pathway from that of β oxidation. Fatty acid synthase is the enzyme complex responsible for fatty acid synthase. Synthesis starts with the carboxylation of acetyl CoA to malonyl CoA, the committed step. This ATP-driven reaction is catalyzed by acetyl CoA carboxylase, a biotin enzyme. The intermediates in fatty acid synthesis are linked to an acyl carrier protein. Acetyl ACP is formed from acetyl CoA, and malonyl ACP is formed from malonyl CoA. Acetyl ACP and malonyl ACP condense to form acetoacetyl ACP, a reaction driven by the release of CO_2 from the activated malonyl unit. A reduction, a dehydration, and a second reduction follow. NADPH is the reductant in these steps. The butyryl ACP formed in this way is ready for a second round of elongation, starting with the addition of a two-carbon unit from malonyl ACP. Seven rounds of elongation yield palmitoyl ACP, which is hydrolyzed to palmitate. In higher organisms, the enzymes catalyzing fatty acid synthesis are covalently linked in a multifunctional enzyme complex. A reaction cycle based on the formation and cleavage of citrate carries acetyl groups from mitochondria to the cytoplasm. NADPH needed for synthesis is generated in the transfer of reducing equivalents from mitochondria by the combined action of cytoplasmic malate dehydrogenase and malic enzyme and by the pentose phosphate pathway.

22.5 The Elongation and Unsaturation of Fatty Acids Are Accomplished by Accessory Enzyme Systems

Fatty acids are elongated and desaturated by enzyme systems in the endoplasmic reticulum membrane. Desaturation requires NADH and O_2 and is carried out by a complex consisting of a flavoprotein, a cytochrome, and a nonheme iron protein. Mammals lack the enzymes to introduce double bonds distal to C-9, and so they require linoleate and linolenate in their diets.

Arachidonate, an essential precursor of prostaglandins and other signal molecules, is derived from linoleate. This 20:4 polyunsaturated fatty acid is the precursor of several classes of signal molecules—prostaglandins, prostacyclins, thromboxanes, and leukotrienes—that act as messengers and local hormones because of their transience. They are called eicosanoids because they contain 20 carbon atoms. Aspirin (acetylsalicylate), an anti-inflammatory and antithrombotic drug, irreversibly blocks the synthesis of these eicosanoids.

22.6 Acetyl CoA Carboxylase Plays a Key Role in Controlling Fatty Acid Metabolism

Fatty acid synthesis and degradation are reciprocally regulated so that both are not simultaneously active. Acetyl CoA carboxylase, the essential control site, is phosphorylated and inactivated by AMP-activated kinase. The phosphorylation is reversed by a protein phosphatase. Citrate, which signals an abundance of building blocks and energy, partly reverses the inhibition by phosphorylation. Carboxylase activity is stimulated by insulin and inhibited by glucagon and epinephrine. In times of plenty, fatty acyl CoAs do not enter the mitochondrial matrix, because malonyl CoA inhibits carnitine acyltransferase I.

KEY TERMS

triacylglycerol (neutral fat, triglyceride) (p. 643)
lipid droplet (p. 645)
acyl adenylate (p. 649)
carnitine (p. 649)
β-oxidation pathway (p. 650)
vitamin B_{12} (cobalamin) (p. 654)
peroxisome (p. 656)
ketone body (p. 657)
acyl carrier protein (ACP) (p. 661)
fatty acid synthase (p. 661)
malonyl CoA (p. 661)
acetyl CoA carboxylase (p. 662)
ATP-citrate lyase (p. 666)
arachidonate (p. 669)
prostaglandin (p. 669)
eicosanoid (p. 669)
polyketide (p. 670)
AMP-activated protein kinase (AMPK) (p. 671)

PROBLEMS

1. *Match'em.* Match each term with its description.

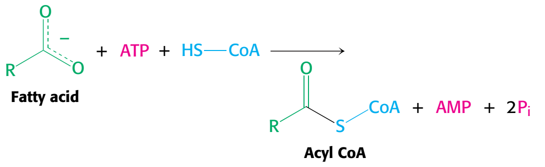
(a) Triacylglycerol ______
(b) Perilipin ______
(c) Adipose triglyceride lipase ______
(d) Glucagon ______
(e) Acyl CoA synthetase ______
(f) Carnitine ______
(g) β-Oxidation pathway ______
(h) Enoyl CoA isomerase ______
(i) 2,4-Dienoyl CoA reductase ______
(j) Methylmalonyl CoA mutase ______
(k) Ketone body ______

1. The enzyme that initiates lipid degradation
2. Activates fatty acids for degradation
3. Converts a *cis*-Δ^3 double bond into a *trans*-Δ^2 double bond
4. Reduces 2,4-dienoyl intermediate to *trans*-Δ^3-enoyl CoA
5. Storage form of fats
6. Required for entry into mitochondria
7. Requires vitamin B_{12}
8. Acetoacetate
9. Means by which fatty acids are degraded
10. Stimulates lipolysis
11. Lipid-droplet-associated protein

2. *After lipolysis.* Write a balanced equation for the conversion of glycerol into pyruvate. Which enzymes are required in addition to those of the glycolytic pathway?

3. *Forms of energy.* The partial reactions leading to the synthesis of acyl CoA (equations 1 and 2, p. 649) are freely reversible. The equilibrium constant for the sum of these reactions is close to 1, meaning that the energy levels of the reactants and products are about equal, even though a molecule of ATP has been hydrolyzed. Explain why these reactions are readily reversible.

4. *Activation fee.* The reaction for the activation of fatty acids before degradation is



(1)

This reaction is quite favorable because the equivalent of two molecules of ATP is hydrolyzed. Explain why, from a biochemical bookkeeping point of view, the equivalent of two molecules of ATP is used despite the fact that the left side of the equation has only one molecule of ATP.

5. *Proper sequence.* Place the following list of reactions or relevant locations in the β oxidation of fatty acids in the proper order.

(a) Reaction with carnitine
(b) Fatty acid in the cytoplasm
(c) Activation of fatty acid by joining to CoA
(d) Hydration
(e) NAD^+-linked oxidation
(f) Thiolysis
(g) Acyl CoA in mitochondrion
(h) FAD-linked oxidation.

6. *Remembrance of reactions past.* We have encountered reactions similar to the oxidation, hydration, and oxidation reactions of fatty acid degradation earlier in our study of biochemistry. What other pathway employs this set of reactions?

7. *A phantom acetyl CoA?* In the equation for fatty acid degradation shown here, only seven molecules of CoA are required to yield eight molecules of acetyl CoA. How is this difference possible?

$$\begin{aligned}\text{Palmitoyl CoA} + 7\,\text{FAD} + 7\,\text{NAD}^+ \\ + 7\,\text{CoA} + 7\,\text{H}_2\text{O} \longrightarrow \\ 8\,\text{acetyl CoA} + 7\,\text{FADH}_2 + 7\,\text{NADH} + 7\,\text{H}^+\end{aligned}$$

8. *Comparing yields.* Compare the ATP yields from palmitic acid and palmitoleic acid.

9. *Counting ATPs 1.* What is the ATP yield for the complete oxidation of C_{17} (heptadecanoic) fatty acid? Assume that the propionyl CoA ultimately yields oxaloacetate in the citric acid cycle.

10. *Sweet temptation.* Stearic acid is a C_{18} fatty acid component of chocolate. Suppose you had a depressing day and decided to settle matters by gorging on chocolate. How much ATP would you derive from the complete oxidation of stearic acid to CO_2?

11. *The best storage form.* Compare the ATP yield from the complete oxidation of glucose, a six-carbon carbohydrate, and hexanoic acid, a six-carbon fatty acid. Hexanoic acid is also called *caprioic acid* and is responsible for the "aroma" of goats. Why are fats better fuels than carbohydrates?

12. *From fatty acid to ketone body.* Write a balanced equation for the conversion of stearate into acetoacetate.

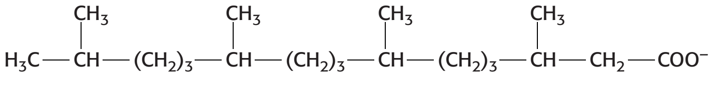
13. *Generous, but not to a fault.* Liver is the primary site of ketone-body synthesis. However, ketone bodies are not used by the liver but are released for other tissues to use. The liver does gain energy in the process of synthesizing and releasing ketone bodies. Calculate the number of molecules of ATP generated by the liver in the conversion of palmitate, a C_{16} fatty acid, into acetoacetate.

14. *Counting ATPs 2.* How much energy is attained with the complete oxidation of the ketone body D-3-hydroxybutyrate?

15. *Another view.* Why might someone argue that the answer to problem 14 is wrong?

16. *An accurate adage.* An old biochemistry adage is that *fats burn in the flame of carbohydrates.* What is the molecular basis of this adage?

17. *Refsum disease.* Phytanic acid is a branched-chain fatty acid component of chlorophyll and is a significant component of milk. In susceptible people, phytanic acid can accumulate, leading to neurological problems. This syndrome is called Refsum disease or phytanic acid storage disease.



$$\mathrm{H_3C-\overset{\overset{\displaystyle CH_3}{|}}{C}H-(CH_2)_3-\overset{\overset{\displaystyle CH_3}{|}}{C}H-(CH_2)_3-\overset{\overset{\displaystyle CH_3}{|}}{C}H-(CH_2)_3-\overset{\overset{\displaystyle CH_3}{|}}{C}H-CH_2-COO^-}$$

Phytanic acid

(a) Why does phytanic acid accumulate?

(b) What enzyme activity would you invent to prevent its accumulation?

18. *A hot diet.* Tritium is a radioactive isotope of hydrogen and can be readily detected. A fully tritiated, six-carbon saturated fatty acid is administered to a rat, and a muscle biopsy of the rat is taken by concerned, sensitive, and discrete technical assistants. These assistants carefully isolate all of the acetyl CoA generated from the β oxidation of the radioactive fatty acid and remove the CoA to form acetate. What will be the overall tritium-to-carbon ratio of the isolated acetate?

19. *Finding triacylglycerols in all the wrong places.* Insulin-dependent diabetes is often accompanied by high levels of triacylglycerols in the blood. Suggest a biochemical explanation.

20. *Match 'em.* Match the terms with the descriptions.

(a) ATP-citrate lyase ____
(b) Malic enzyme ____
(c) Malonyl CoA ____
(d) Acetyl CoA carboxylase ____
(e) Acyl carrier protein ____
(f) β-Ketoacyl synthase ____
(g) Palmitate ____
(h) Eicosanoids ____
(i) Arachidonate ____
(j) AMP-activated protein kinase ____

1. Helps to generate NADPH from NADH
2. Inactivates acetyl CoA carboxylase
3. Molecule on which fatty acids are synthesized
4. A precursor of prostaglandins
5. Activated acetyl CoA
6. The end product of fatty acid synthase
7. Fatty acids containing 20 carbon atoms
8. Catalyzes the committed step in fatty acid synthesis
9. Catalyzes the reaction of acetyl ACP and malonyl ACP
10. Generates cytoplasmic acetyl CoA

21. *Counterpoint.* Compare and contrast fatty acid oxidation and synthesis with respect to

(a) site of the process

(b) acyl carrier

(c) reductants and oxidants

(d) stereochemistry of the intermediates

(e) direction of synthesis or degradation

(f) organization of the enzyme system

22. *The fizz in fatty acid synthesis.* During the initial in vitro studies on fatty acid synthesis, experiments were performed to determine which buffer would result in optimal activity. Bicarbonate buffer proved far superior to phosphate buffer. The researchers could not initially account for this result. From your perspective of many decades later, explain why the result is not surprising.

23. *A supple synthesis.* Myristate, a saturated C_{14} fatty acid, is used as an emollient for cosmetics and topical medicinal preparations. Write a balanced equation for the synthesis of myristate.

24. *The cost of cleanliness.* Lauric acid is a 12-carbon fatty acid with no double bonds. The sodium salt of lauric acid (sodium laurate) is a common detergent used in a variety of products, including laundry detergent, shampoo, and toothpaste. How many molecules of ATP and NADPH are required to synthesize lauric acid?

25. *Proper organization.* Arrange the following steps in fatty acid synthesis in their proper order.

(a) Dehydration
(b) Condensation
(c) Release of a C_{16} fatty acid
(d) Reduction of a carbonyl group
(e) Formation of malonyl ACP

26. *No access to assets.* What would be the effect on fatty acid synthesis of a mutation in ATP-citrate lyase that reduces the enzyme's activity? Explain.

27. *The truth and nothing but.* True or False. If false, explain.

(a) Biotin is required for fatty acid synthase activity.

(b) The condensation reaction in fatty acid synthesis is powered by the decarboxylation of malonyl CoA.

(c) Fatty acid synthesis does not depend on ATP.

(d) Palmitate is the end product of fatty acid synthase.

(e) All of the enzyme activities required for fatty acid synthesis in mammals are contained in a single polypeptide chain.

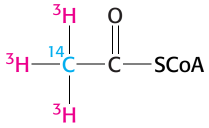
(f) Fatty acid synthase in mammals is active as a monomer.

(g) The fatty acid arachidonate is a precursor for signal molecules.

(h) Acetyl CoA carboxylase is inhibited by citrate.

28. *Odd fat out.* Suggest how fatty acids with odd numbers of carbons are synthesized.

29. *Labels.* Suppose that you had an in vitro fatty acid-synthesizing system that had all of the enzymes and cofactors required for fatty acid synthesis except for acetyl CoA. To this system, you added acetyl CoA that contained radioactive hydrogen (^{3}H, tritium) and carbon 14 (^{14}C) as shown here.



$$ {}^{3}\mathrm{H}{-}\overset{{}^{3}\mathrm{H}}{\underset{{}^{3}\mathrm{H}}{{}^{14}\mathrm{C}}}{-}\overset{\mathrm{O}}{\overset{\|}{\mathrm{C}}}{-}\mathrm{SCoA} $$

The ratio of $^{3}H/^{14}C$ is 3. What would the $^{3}H/^{14}C$ ratio be after the synthesis of palmitic acid (C_{16}) with the use of the radioactive acetyl CoA?

30. *A tight embrace.* Avidin, a glycoprotein found in eggs, has a high affinity for biotin. Avidin can bind biotin and prevent its use by the body. How might a diet rich in raw eggs affect fatty acid synthesis? What will be the effect on fatty acid synthesis of a diet rich in cooked eggs? Explain.

31. *Alpha or omega?* Only one acetyl CoA molecule is used directly in fatty acid synthesis. Identify the carbon atoms in palmitic acid that were donated by acetyl CoA.

32. *Now you see it, now you don't.* Although HCO_3^- is required for fatty acid synthesis, its carbon atom does not appear in the product. Explain how this omission is possible.

33. *It is all about communication.* Why is citrate an appropriate inhibitor of phosphofructokinase?

34. *Tracing carbon atoms.* Consider a cell extract that actively synthesizes palmitate. Suppose that a fatty acid synthase in this preparation forms one molecule of palmitate in about 5 minutes. A large amount of malonyl CoA labeled with ^{14}C in each carbon atom of its malonyl unit is suddenly added to this system, and fatty acid synthesis is stopped a minute later by altering the pH. The fatty acids are analyzed for radioactivity. Which carbon atom of the palmitate formed by this system is more radioactive—C-1 or C-14?

35. *An unaccepting mutant.* The serine residues in acetyl CoA carboxylase that are the target of the AMP-activated protein kinase are mutated to alanine. What is a likely consequence of this mutation?

36. *Sources.* For each of the following unsaturated fatty acids, indicate whether the biosynthetic precursor in animals is palmitoleate, oleate, linoleate, or linolenate.

(a) 18:1 *cis*-Δ^{11}
(b) 18:3 *cis*-Δ^{6}, Δ^{9}, Δ^{12}
(c) 20:2 *cis*-Δ^{11}, Δ^{14}
(d) 20:3 *cis*-Δ^{5}, Δ^{8}, Δ^{11}
(e) 22:1 *cis*-Δ^{13}
(f) 22:6 *cis*-Δ^{4}, Δ^{7}, Δ^{10}, Δ^{13}, Δ^{16}, Δ^{19}

37. *Driven by decarboxylation.* What is the role of decarboxylation in fatty acid synthesis? Name another key reaction in a metabolic pathway that employs this mechanistic motif.

38. *Kinase surfeit.* Suppose that a promoter mutation leads to the overproduction of protein kinase A in adipose cells. How might fatty acid metabolism be altered by this mutation?

39. *Blocked assets.* The presence of a fuel molecule in the cytoplasm does not ensure that the fuel molecule can be effectively used. Give two examples of how impaired transport of metabolites between compartments leads to disease.

40. *Elegant inversion.* Peroxisomes have an alternative pathway for oxidizing polyunsaturated fatty acids. They contain a hydratase that converts D-3-hydroxyacyl CoA into *trans*-Δ^{2}-enoyl CoA. How can this enzyme be used to oxidize CoAs containing a *cis* double bond at an even-numbered carbon atom (e.g., the *cis*-Δ^{12} double bond of linoleate)?

41. *Arm in arm.* Many eukaryotic enzymes, such as fatty acid synthase, are multifunctional proteins with multiple active sites covalently linked. What is the advantage of this arrangement?

42 *Covalent catastrophe.* What is a potential disadvantage of having many catalytic sites together on one very long polypeptide chain?

43. *Missing acyl CoA dehydrogenases.* A number of genetic deficiencies in acyl CoA dehydrogenases have been described. This deficiency presents early in life after a period of fasting. Symptoms include vomiting, lethargy, and sometimes coma. Not only are blood levels of glucose low (hypoglycemia), but starvation-induced ketosis is absent. Provide a biochemical explanation for these last two observations.

44. *Effects of clofibrate.* High blood levels of triacylglycerides are associated with heart attacks and strokes. Clofibrate, a drug that increases the activity of peroxisomes, is sometimes used to treat patients with such a condition. What is the biochemical basis for this treatment?

45. *A different kind of enzyme.* Figure 22.36 shows the response of acetyl CoA carboxylase to varying amounts of citrate. Explain this effect in light of the allosteric effects that citrate has on the enzyme. Predict the effects of increasing concentrations of palmitoyl CoA.

Mechanism Problems

46. *Variation on a theme.* Thiolase is homologous in structure to the condensing enzyme. On the basis of this observation, propose a mechanism for the cleavage of 3-ketoacyl CoA by CoA.

47. *Two plus three to make four.* Propose a reaction mechanism for the condensation of an acetyl unit with a malonyl unit to form an acetoacetyl unit in fatty acid synthesis.

Chapter Integration Problems

48. *Ill-advised diet.* Suppose that, for some bizarre reason, you decided to exist on a diet of whale and seal blubber, exclusively.

(a) How would lack of carbohydrates affect your ability to utilize fats?

(b) What would your breath smell like?

(c) One of your best friends, after trying unsuccessfully to convince you to abandon this diet, makes you promise to consume a healthy dose of odd-chain fatty acids. Does your friend have your best interests at heart? Explain.

49. *Fats to glycogen.* An animal is fed stearic acid that is radioactively labeled with [^{14}C]carbon. A liver biopsy reveals the presence of ^{14}C-labeled glycogen. How is this finding possible in light of the fact that animals cannot convert fats into carbohydrates?

50. *Enzyme deficit.* Pyruvate dehydrogenase deficiency, a severe condition with a poor prognosis, results in neurological dysfunction and unremitting lactic acidosis. One treatment is placing the patients on a low carbohydrate/high fat diet.

(a) Explain why the deficiency would be characterized by neurological dysfunction.

(b) Why does lactic acidosis result?

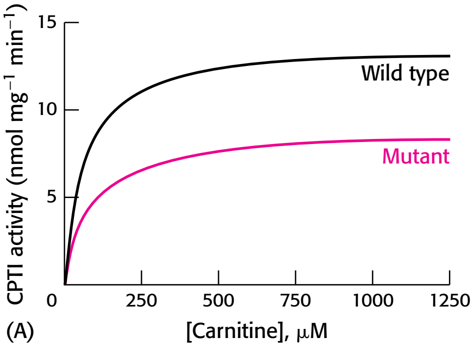
(c) Explain the rationale for the low carbohydrate/high fat diet.

51. *Right enzymes. Wrong location.* Zwellweger syndrome is caused by a defect in enzyme importation into the peroxisomes. What disease of carbohydrate metabolism is caused by inappropriate cellular distribution of enzymes?

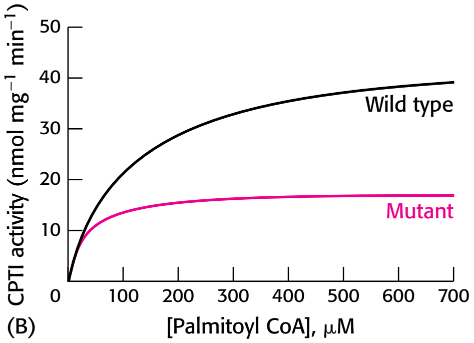
Data Interpretation Problem

52. *Mutant enzyme.* Carnitine palmitoyl transferase I (CPTI) catalyzes the conversion of long-chain acyl CoA into acyl carnitine, a prerequisite for transport into mitochondria and subsequent degradation. A mutant enzyme was constructed with a single amino acid change at position 3 of glutamic acid for alanine. Graphs A through C show data from studies performed to identify the effect of the mutation.

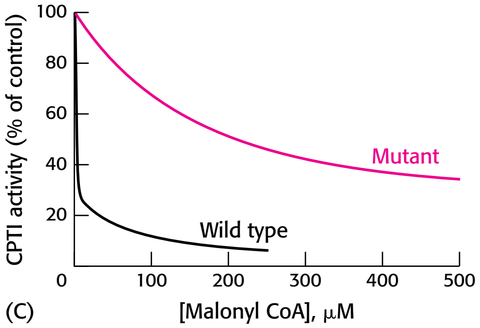
(a) What is the effect of the mutation on enzyme activity when the concentration of carnitine is varied (Graph A)? What are the K_M and V_{max} values for the wild-type and mutant enzymes?



(b) What is the effect when the experiment is repeated with varying concentrations of palmitoyl CoA (Graph B)? What are the K_M and V_{max} values for the wild-type and mutant enzymes?



(c) Graph C shows the inhibitory effect of malonyl CoA on the wild-type and mutant enzymes. Which enzyme is more sensitive to malonyl CoA inhibition?



(d) Suppose that the concentration of palmitoyl CoA = 100 μM, that of carnitine = 100 μM, and that of malonyl CoA = 10 μM. Under these conditions, what is the most prominent effect of the mutation on the properties of the enzyme?

(e) What can you conclude about the role of glutamate 3 in CPTI function?

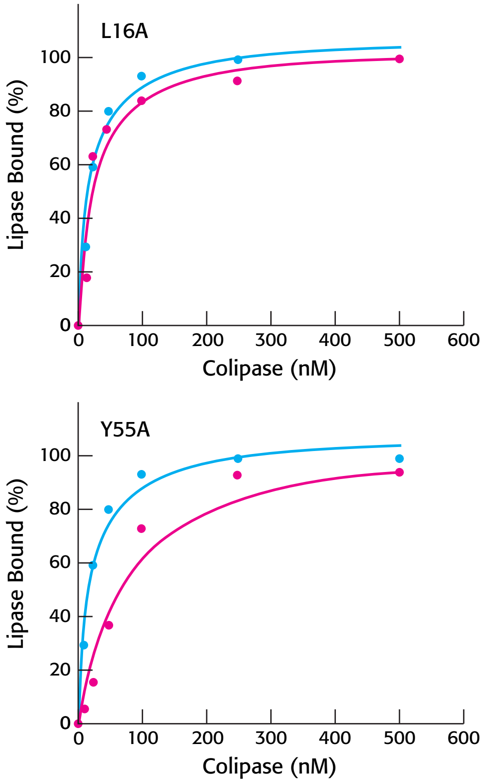
[Data from J. Shi, H. Zhu, D. N. Arvidson, and G. J. Woldegiorgis. *J. Biol. Chem.* 274:9421–9426, 1999.]

53. *Partners*. Colipase binds to lipid particles in the intestine and facilitates the subsequent binding and activation of pancreatic lipase. The lipase then hydrolyzes the lipids so that they can be absorbed by intestinal cells. Experiments were undertaken using site-directed mutagenesis (Section 5.2) to determine which amino acids in the colipase facilitated lipase binding and which activated the lipase. A portion of these experiments is presented below.

Figure A shows the results of two different amino acid substitutions on the ability of colipase to facilitate the binding of lipase to the lipid particles. In both panels, the blue circles show the control (no amino acid changes) data and the red circles the data with the altered amino acid. The first letter represents the normal amino acid, the number its position in the amino acid sequence, and the second letter the replacement amino acid.

(a) What is the amino acid alteration in the upper panel?

(A)



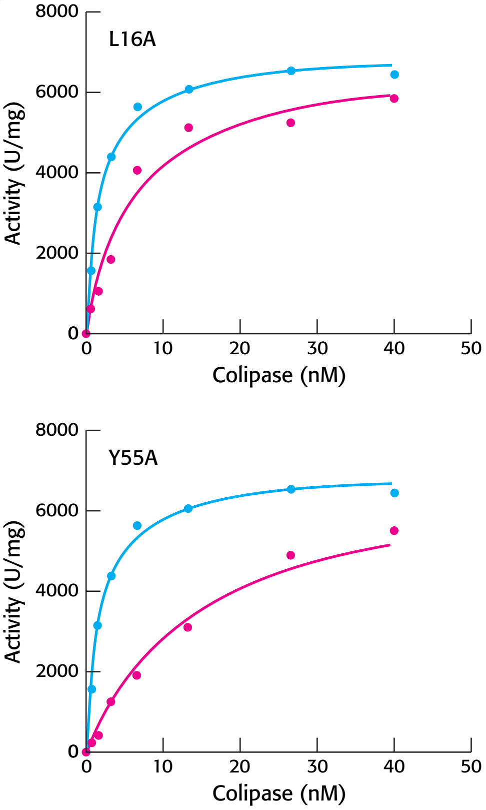
(b) How does this mutation change the ability of colipase to bind the lipase to the lipid particle?

(c) What is the amino acid substitution in the lower panel?

(d) What is the effect of this substitution on the ability of the colipase to bind the lipase to the lipid particle?

Next, the activity of the bound lipase was examined, as shown in Figure B.

(B)



[Data from: Ross, L. E., Xiao, X., and Lowe, M. E. 2013. Identification of amino acids in human colipase that mediate adsorption to lipid emulsions and mixed micelles. *Biochim. Biophys. Acta* 1831: 1052–1059.]

(e) What was the effect of the L16A mutation on lipase activity?

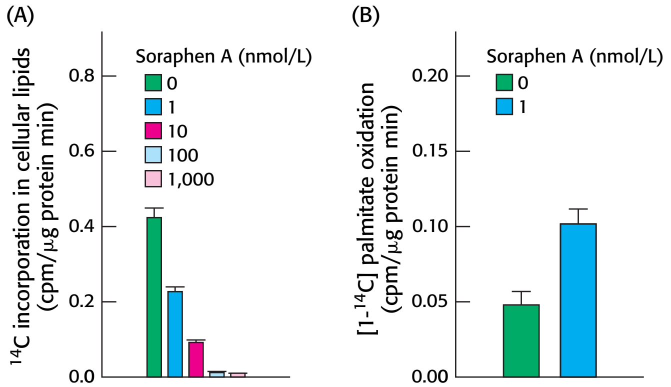
(f) What does this suggest about the lipase binding and activation properties of the colipase?

(g) What was the effect of the Y55A mutation on lipase activity?

54. *Nip it in the bud.* Soraphen A is a natural antifungal agent isolated from a myxobacterium. Soraphen A is also a potent inhibitor of acetyl CoA carboxylase, the regulatory enzyme that initiates fatty acid synthesis. As discussed earlier (p. 667), cancer cells must produce large amount of fatty acids to generate phospholipids for membrane synthesis. Thus, acetyl CoA carboxylase might be a target for anti-cancer drugs.

Following are the results of experiments testing this hypothesis on a cancer line.

Figure A shows the effect of various amounts of soraphen A on fatty acid synthesis as measured by the incorporation of radioactive acetate (^{14}C) into fatty acids.



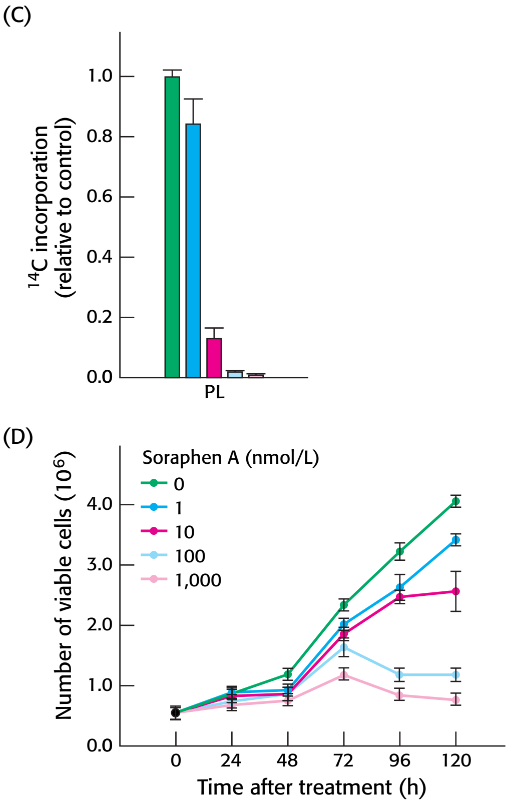
(a) How did soraphen A alter fatty acid synthesis? How was synthesis affected by increasing concentrations of soraphen A?

Figure B shows the results obtained when fatty acid oxidation was measured, as the release of radioactive CO_2 from added radioactive (^{14}C) palmitate, in the presence of soraphen A.

(b) How did soraphen A alter fatty acid oxidation?

(c) Explain the results obtained in B in light of the fact that soraphen A inhibits acetyl CoA carboxylase.

More experiments were undertaken to assess whether carboxylase inhibition did indeed prevent phospholipid synthesis. Again, cells were grown in the presence of radioactive acetate and phospholipids were subsequently isolated and the amount of ^{14}C incorporated into the phospholipid was determined. The effect of soraphen A on phospholipid synthesis is shown in Figure C.



[Data from: Beckers, A., Organe, S., Timmermans, L., Scheys, K., Peeters, A., Brusselmans, K., Verhoeven, G., and Swinnen, J. V. 2007. Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells. *Cancer Res*. 67: 8180–8187.]

(d) Did the inhibition of carboxylase by the drug alter phospholipid synthesis?

(e) How might phospholipid synthesis affect cell viability?

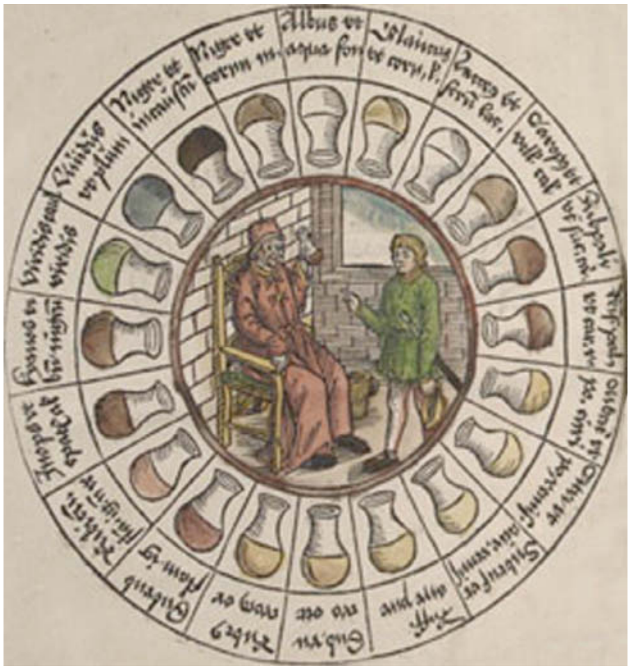
Finally, as shown in Figure D, experiments were performed to determine if the drug inhibits cancer growth in vitro by determining the number of cancer cells surviving upon exposure to soraphen A.

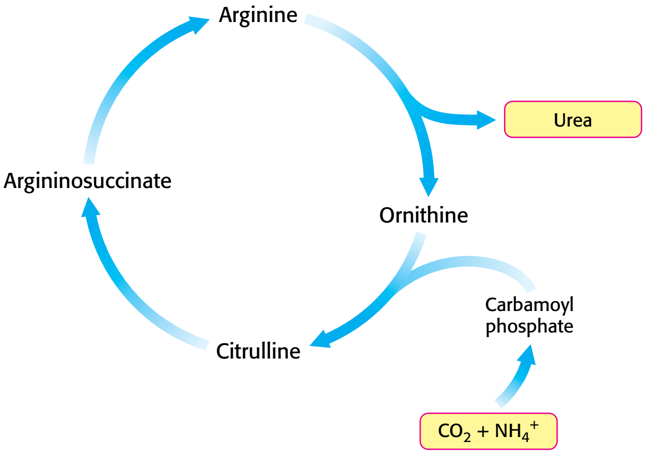
(f) How did soraphen A affect cancer cell viability?

CHAPTER

23

Protein Turnover and Amino Acid Catabolism





This fourteenth-century hand-colored woodcut from Germany depicts a wheel that classifies urine samples according to their color and consistency. In the middle of the wheel, a doctor inspects a patient's urine by sight, smell, and taste. The vials on the wheel aided physicians in diagnosing diseases. A key component of urine is urea, which is formed from the amino groups released during the metabolism of amino acids. [(Left) Rosenwald Collection. Rare Book and Special Collections Division, Library of Congress (128.2).]

The digestion of dietary proteins in the intestine and the degradation of proteins within the cell provide a steady supply of amino acids to the cell. Many cellular proteins are constantly degraded and resynthesized in response to changing metabolic demands. Others are misfolded or become damaged and they, too, must be degraded. Unneeded or damaged proteins are marked for destruction by the covalent attachment of chains of a small protein called *ubiquitin* and then degraded by a large, ATP-dependent complex called the *proteasome. The primary use of amino acids provided through degradation or digestion is as building blocks for the synthesis of proteins and other nitrogenous compounds such as nucleotide bases.*

Amino acids in excess of those needed for biosynthesis can neither be stored, in contrast with fatty acids and glucose, nor excreted. Rather, surplus amino acids are used as metabolic fuel. *The α-amino group is removed, and the resulting carbon skeleton is converted into a major metabolic intermediate.* Most of the amino groups harvested from surplus amino acids are converted into urea through the *urea cycle,* and their carbon skeletons are transformed into acetyl CoA, acetoacetyl CoA, pyruvate, or one of the intermediates of the citric acid cycle. The carbon skeletons are converted into glucose, glycogen, and fats.

OUTLINE

23.1 Proteins Are Degraded to Amino Acids
23.2 Protein Turnover Is Tightly Regulated
23.3 The First Step in Amino Acid Degradation Is the Removal of Nitrogen
23.4 Ammonium Ion Is Converted into Urea in Most Terrestrial Vertebrates
23.5 Carbon Atoms of Degraded Amino Acids Emerge as Major Metabolic Intermediates
23.6 Inborn Errors of Metabolism Can Disrupt Amino Acid Degradation

Several coenzymes play key roles in amino acid degradation; foremost among them is *pyridoxal phosphate.* This coenzyme forms Schiff-base intermediates, which are a type of aldimine, that allow α-amino groups to be shuttled between amino acids and ketoacids. We will consider several genetic errors of amino acid degradation that lead to brain damage and mental retardation unless remedial action is initiated soon after birth. The study of amino acid metabolism is especially rewarding because it is rich in connections between basic biochemistry and clinical medicine.

23.1 Proteins Are Degraded to Amino Acids

Dietary protein is a vital source of amino acids. Especially important dietary proteins are those containing the essential amino acids—amino acids that cannot be synthesized and must be acquired in the diet (Table 23.1). Proteins ingested in the diet are digested into amino acids or small peptides that can be absorbed by the intestine and transported in the blood. Another crucial source of amino acids is the degradation of cellular proteins.

TABLE 23.1 Essential amino acids in human beings

Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Threonine
Tryptophan
Valine

The digestion of dietary proteins begins in the stomach and is completed in the intestine

Protein digestion begins in the stomach, where the acidic environment denatures proteins into random coils. Denatured proteins are more accessible as substrates for proteolysis than are native proteins. The primary proteolytic enzyme of the stomach is *pepsin,* a nonspecific protease that, remarkably, is maximally active at pH 2. Thus, pepsin can function in the highly acidic environment of the stomach that disables other proteins.

The partly digested proteins then move from the acidic environment of the stomach to the beginning of the small intestine. The low pH of the food as well as the polypeptide products of pepsin digestion stimulate the release of hormones that promote the secretion from the pancreas of sodium bicarbonate ($NaHCO_3$), which neutralizes the pH of the food, and a variety of pancreatic proteolytic enzymes. Recall that these enzymes are secreted as inactive zymogens that are then converted into active enzymes (Sections 9.1 and 10.4). The battery of enzymes displays a wide array of specificity, and so the substrates are degraded into free amino acids as well as di- and tripeptides. Digestion is further enhanced by proteolytic enzymes, such as aminopeptidase N, that are located in the plasma membrane of the intestinal cells. Aminopeptidases digest proteins from the amino-terminal end. Single amino acids, as well as di- and tripeptides, are transported into the intestinal cells.

At least seven different transporters exist, each specific to a different group of amino acids. A number of inherited disorders result from mutations in these transporters. For example, Hartnup disease, a rare disorder characterized by rashes, ataxia (lack of muscle control), delayed mental development, and diarrhea, results from a defect in the transporter for tryptophan and other nonpolar amino acids. The absorbed amino acids are subsequently released into the blood by a number of Na^+–amino acid antiporters for use by other tissues (Figure 23.1).

Cellular proteins are degraded at different rates

Protein turnover—the degradation and resynthesis of proteins—takes place constantly in cells. Although some proteins are very stable, many proteins are short lived, particularly those that participate in metabolic regulation. These proteins can be quickly degraded to activate or shut down a signaling

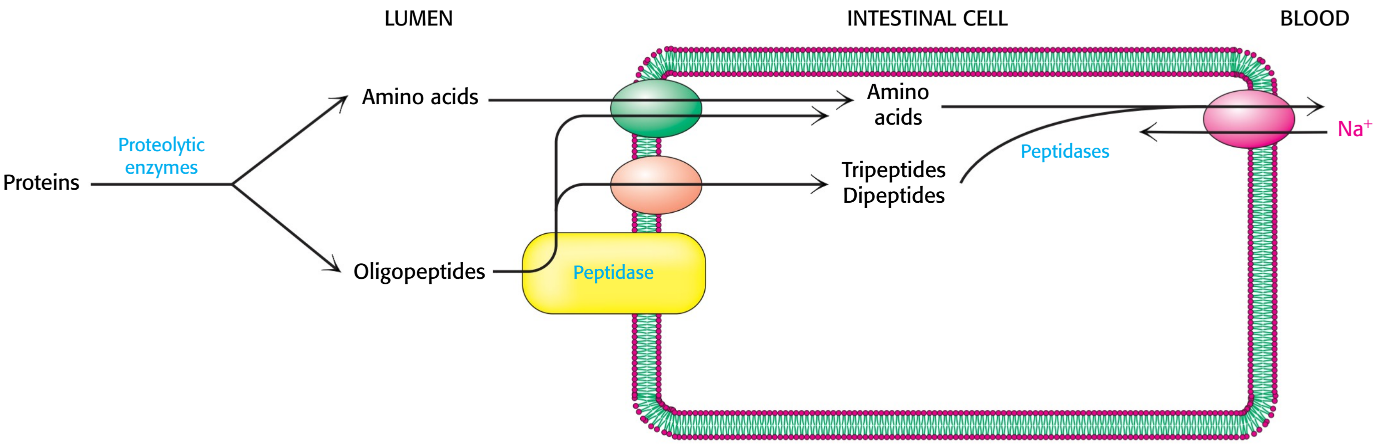


FIGURE 23.1 Digestion and absorption of proteins. Protein digestion is primarily a result of the activity of enzymes secreted by the pancreas. Aminopeptidases associated with the intestinal epithelium further digest proteins. The amino acids and di- and tripeptides are absorbed into the intestinal cells by specific transporters (green and orange ovals). Free amino acids are then released into the blood by transporters (red oval) for use by other tissues.

pathway. In addition, cells must eliminate damaged proteins. A significant proportion of newly synthesized protein molecules are defective because of errors in translation or misfolding. Even proteins that are normal when first synthesized may undergo oxidative damage or be altered in other ways with the passage of time. These proteins must be removed before they accumulate and aggregate. Indeed, a number of pathological conditions, such as certain forms of Parkinson disease and Huntington disease, are associated with protein aggregation.

The half-lives of proteins range over several orders of magnitude. Ornithine decarboxylase, at approximately 11 minutes, has one of the shortest half-lives of any mammalian protein. This enzyme participates in the synthesis of polyamines, which are cellular cations essential for growth and differentiation. The life of hemoglobin, however, is limited only by the life of the red blood cell, and the lens protein crystallin is limited by the life of the organism.

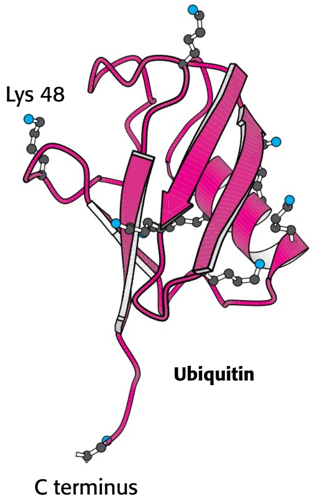


FIGURE 23.2 Structure of ubiquitin. *Notice* that ubiquitin has an extended carboxyl terminus, which is activated and linked to proteins targeted for destruction. Lysine residues, including lysine 48, the major site for linking additional ubiquitin molecules, are shown as ball-and-stick models. [Drawn from 1UBI.pdb.]

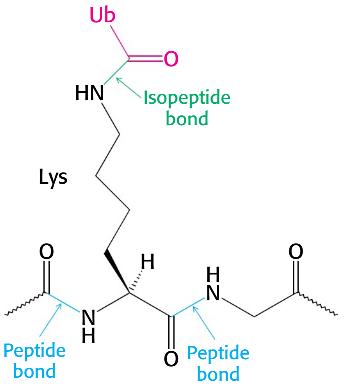
23.2 Protein Turnover Is Tightly Regulated

How can a cell distinguish proteins that should be degraded? *Ubiquitin* (Ub), a small (76 amino acids) protein present in all eukaryotic cells, is a tag that marks proteins for destruction (Figure 23.2). Ubiquitin is the cellular equivalent of the "black spot" of Robert Louis Stevenson's *Treasure Island*—the signal for death.

Ubiquitin tags proteins for destruction

Ubiquitin is highly conserved in eukaryotes: yeast and human ubiquitin differ at only 3 of 76 residues. The carboxyl-terminal glycine residue of ubiquitin becomes covalently attached to the ε-amino groups of several lysine residues on a protein destined to be degraded. The energy for the formation of these *isopeptide bonds* (*iso* because ε- rather than α-amino groups are targeted) comes from ATP hydrolysis.

Three enzymes participate in the attachment of ubiquitin to a protein (Figure 23.3): ubiquitin-activating enzyme, or E1; ubiquitin-conjugating enzyme, or E2; and ubiquitin–protein ligase, or E3. First, the C-terminal carboxylate group of ubiquitin becomes linked to a sulfhydryl group of E1 by a thioester bond. This ATP-driven reaction is reminiscent of fatty acid activation (Section 22.2). In this reaction, an acyl adenylate is formed at the



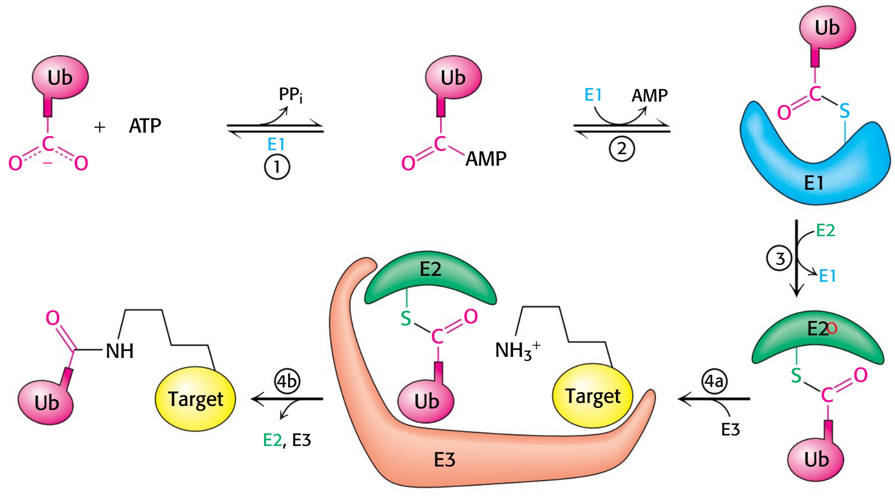


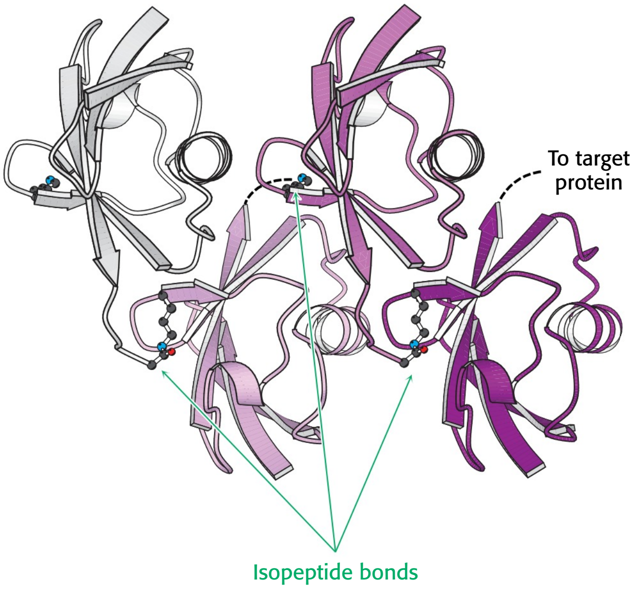
FIGURE 23.3 Ubiquitin conjugation. The ubiquitin-activating enzyme E1 adenylates ubiquitin (Ub) (1) and transfers the ubiquitin to one of its own cysteine residues (2). Ubiquitin is then transferred to a cysteine residue in the ubiquitin-conjugating enzyme E2 by the E2 enzyme. (3). Finally, the ubiquitin–protein ligase E3 transfers the ubiquitin to a lysine residue on the target protein (4a and 4b).

C-terminal carboxylate of ubiquitin with the release of pyrophosphate, and the ubiquitin is subsequently transferred to a sulfhydryl group of a key cysteine residue in E1. The activated ubiquitin is then shuttled to a sulfhydryl group of E2, a reaction catalyzed by E2 itself. Finally, E3 catalyzes the transfer of ubiquitin from E2 to an ε-amino group on the target protein. The ubiquitination reaction is processive: E3 remains bound to the target proteins and generates a chain of ubiquitin molecules by linking the ε-amino group of lysine residue 48 of one ubiquitin molecule to the terminal carboxylate of another. A chain of four or more ubiquitin molecules is especially effective in signaling the need for degradation (Figure 23.4).

What determines whether a protein becomes ubiquitinated? A specific sequence of amino acids, termed a *degron,* indicates that a protein should be degraded. One such signal appears to be quite simple. *The half-life of a cytoplasmic protein is determined to a large extent by its amino-terminal residue* (Table 23.2). This dependency is referred to as the *N-terminal rule* or the *N-terminal degron*. A yeast protein with methionine at its N terminus typically has a half-life of more than 20 hours, whereas one with arginine at this position has a half-life of about 2 minutes. A highly destabilizing N-terminal residue such as arginine or leucine favors rapid ubiquitination, whereas a stabilizing residue such as methionine or proline does not. Why is the N-terminal rule only *apparently* simple? Sometimes the N-terminal degron is exposed only after the protein is proteolytically cleaved. Such degrons are called *pro-N-terminal degrons,* in analogy to proenzymes (Section 10.4), because the protein must be cleaved to expose the signal. In other cases, the destabilizing amino acid is added to the protein after the protein is synthesized. Other modifications, notably N-terminal acetylation, can also activate a degron. Additional degrons thought to identify proteins for degradation include *cyclin destruction boxes,* which are amino acid sequences that mark cell-cycle proteins for destruction, and *PEST sequences*, which contain the amino acid sequence proline (P, single-letter abbreviation), glutamic acid (E), serine (S), and threonine (T).

E3 enzymes are the readers of N-terminal residues. Although most eukaryotes have only one or a small number of distinct E1 enzymes, all eukaryotes have many distinct E2 and E3 enzymes. Moreover, there appears to be only a single family of evolutionarily related E2 proteins but three distinct families of E3 proteins, all together consisting of

FIGURE 23.4 Structure of tetraubiquitin. Four ubiquitin molecules are linked by isopeptide bonds. *Notice* that each isopeptide bond is formed by the linkage of the carboxylate group at the end of the extended C terminus with the ε-amino group of a lysine residue. Dashed lines indicate the positions of the extended C-termini that were not observed in the crystal structure. This unit is the primary signal for degradation when linked to a target protein. [Drawn from 1TBE. pdb.]



hundreds of members. Indeed, the E3 family is one of the largest gene families in human beings. The diversity of target proteins that must be tagged for destruction requires a large number of E3 proteins as readers.

TABLE 23.2 Dependence of the half-lives of cytoplasmic yeast proteins on the identity of their amino-terminal residues

Highly stabilizing residues ($t_{1/2} > 20$ hours)			
Ala	Cys	Gly	Met
Pro	Ser	Thr	Val
Intrinsically destabilizing residues ($t_{1/2} = 2$ to 30 minutes)			
Arg	His	Ile	Leu
Lys	Phe	Trp	Tyr
Destabilizing residues after chemical modification ($t_{1/2} = 3$ to 30 minutes)			
Asn	Asp	Gln	Glu

Data from J. W. Tobias, T. E. Schrader, G. Rocap, and A. Varshavsky. *Science* 254(1991):1374–1377.

Three examples demonstrate the importance of E3 proteins to normal cell function. Proteins that are not broken down owing to a defective E3 may accumulate to create a disease of protein aggregation such as juvenile or early-onset Parkinson disease. A defect in another member of the E3 family causes Angelman syndrome, a severe neurological disorder characterized by mental retardation, absence of speech, uncoordinated movement, and hyperactivity. Conversely, uncontrolled protein turnover also can create dangerous pathological conditions. For example, human papillomavirus (HPV) encodes a protein that activates a specific E3 enzyme. The enzyme ubiquitinates the tumor suppressor p53 and other proteins that control DNA repair, which are then destroyed. The activation of this E3 enzyme is observed in more than 90% of cervical carcinomas. Thus, the inappropriate marking of key regulatory proteins for destruction can trigger further events, leading to tumor formation.

It is important to note that the role of ubiquitin is much broader than merely marking proteins for destruction. Although we have focused on protein degradation, monoubiquitination also regulates proteins involved in DNA repair, chromatin remodeling, and protein kinase activation, among other biochemical processes.

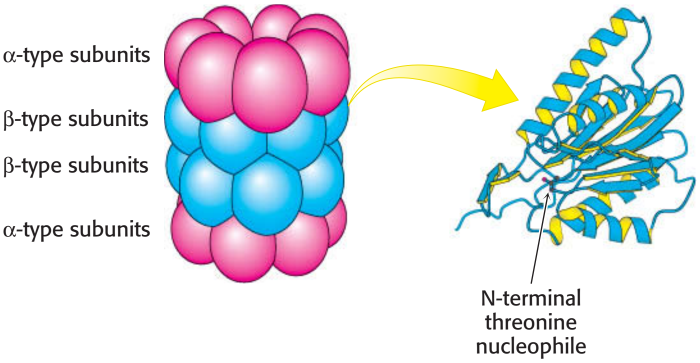


FIGURE 23.5 20S proteasome. The 20S proteasome comprises 28 homologous subunits (α-type, red; β-type, blue), arranged in four rings of 7 subunits each. Some of the β-type subunits (right) include protease active sites at their amino termini. [Subunit drawn from 1RYP.pdb.]

The proteasome digests the ubiquitin-tagged proteins

If ubiquitin is the mark of death, what is the executioner? *A large protease complex called the proteasome or the 26S proteasome digests the ubiquitinated proteins.* This ATP-driven multisubunit protease spares ubiquitin, which is then recycled. The 26S proteasome is a complex of two components: a 20S catalytic unit and a 19S regulatory unit.

The 20S unit is composed of 28 subunits, encoded by 14 genes, arranged into 4 heteroheptomeric rings stacked to form a structure resembling a barrel (Figure 23.5). The outer two rings of the barrel are made up of α-type subunits and the inner two rings of β-type subunits. The 20S catalytic core is a sealed barrel. *Access to its interior is controlled by a 19S regulatory unit,* itself a 700-kDa complex made up of 19 subunits. Two such 19S complexes bind to the 20S proteasome core, one at each end, to form the complete 26S proteasome (Figure 23.6). The 19S regulatory unit has three functions. First, components of the 19S unit are ubiquitin receptors that bind specifically to polyubiquitin chains, thereby ensuring that only ubiquitinated proteins are degraded. Second, an isopeptidase in the 19S unit cleaves off intact ubiquitin molecules from the proteins so that they can be reused. Finally, the doomed protein is unfolded and directed into the catalytic core. Key components of the 19S complex are six ATPases of a type called the AAA class (*A*TPase *a*ssociated with various cellular *a*ctivities). ATP hydrolysis assists the 19S complex to unfold the substrate and induce conformational changes in the 20S catalytic core so that the substrate can be passed into the center of the complex.

The proteolytic active sites are sequestered in the interior of the barrel to protect potential substrates until they are directed into the barrel. There are three types of active sites in the β subunits, each with a different specificity,

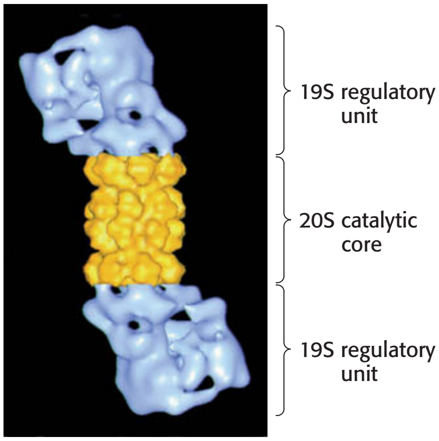
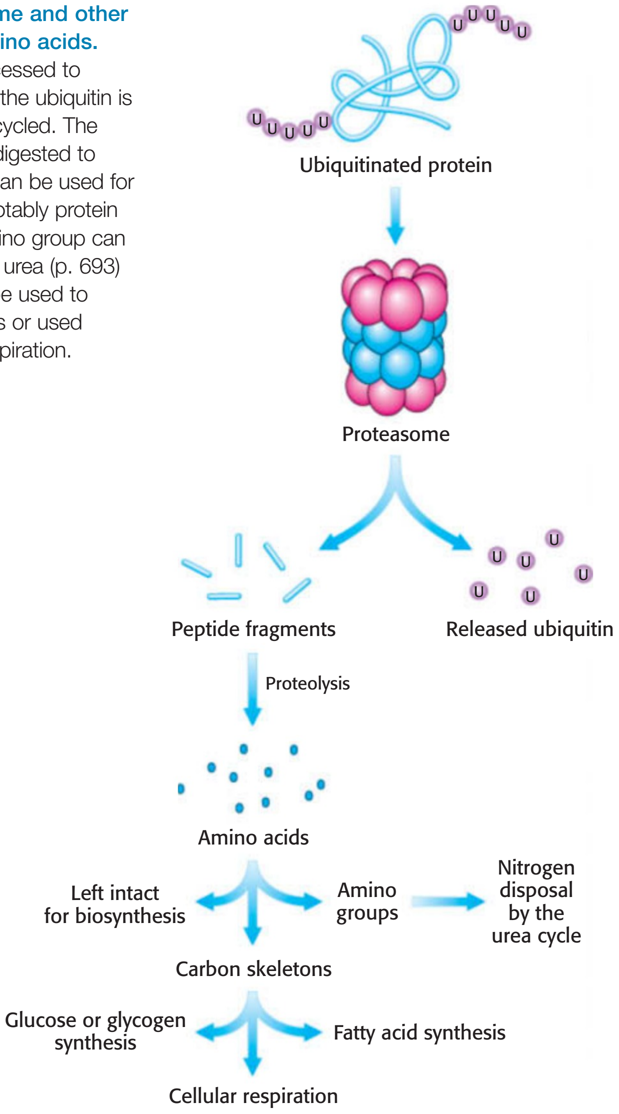


FIGURE 23.6 26S proteasome. A 19S regulatory subunit is attached to each end of the 20S catalytic unit. [From W. Baumeister, J. Walz, F. Zuhl, and E. Seemuller. *Cell* 92(1998):367–380; courtesy of Dr. Wolfgang Baumeister.]

FIGURE 23.7 The proteasome and other proteases generate free amino acids. Ubiquitinated proteins are processed to peptide fragments from which the ubiquitin is subsequently removed and recycled. The peptide fragments are further digested to yield free amino acids, which can be used for biosynthetic reactions, most notably protein synthesis. Alternatively, the amino group can be removed and processed to urea (p. 693) and the carbon skeleton can be used to synthesize carbohydrate or fats or used directly as a fuel for cellular respiration.


U U U U U
U U U U U
Ubiquitinated protein
Proteasome
Peptide fragments
Released ubiquitin
U U U U U U
Proteolysis
Amino acids
Left intact for biosynthesis
Amino groups
Nitrogen disposal by the urea cycle
Carbon skeletons
Glucose or glycogen synthesis
Fatty acid synthesis
Cellular respiration

but all employ an N-terminal threonine. The hydroxyl group of the threonine residue is converted into a nucleophile that attacks the carbonyl groups of peptide bonds to form acyl-enzyme intermediates. Substrates are degraded in a processive manner without the release of degradation intermediates, until the substrate is reduced to peptides ranging in length from seven to nine residues. These peptide products are released from the proteasome and further degraded by other cellular proteases to yield individual amino acids. Thus, the ubiquitination pathway and the proteasome cooperate to degrade unwanted proteins. Figure 23.7 presents an overview of the fates of amino acids following proteasomal digestion.

The ubiquitin pathway and the proteasome have prokaryotic counterparts

Both the ubiquitin pathway and the proteasome appear to be present in all eukaryotes. Homologs of the proteasome are also found in some prokaryotes. The proteasomes of some archaea are quite similar in overall structure to their eukaryotic counterparts and similarly have 28 subunits (Figure 23.8). In the archaeal proteasome, however, all α outer-ring subunits and all β inner-ring subunits are identical; in eukaryotes, each α or β subunit is one of seven different isoforms. This specialization provides distinct substrate specificity.

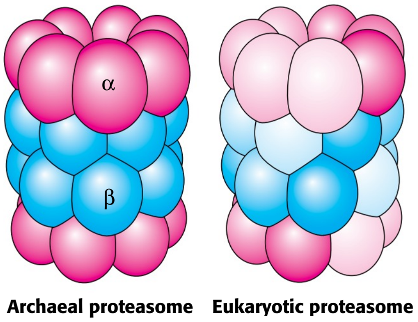

α
β
Archaeal proteasome
Eukaryotic proteasome

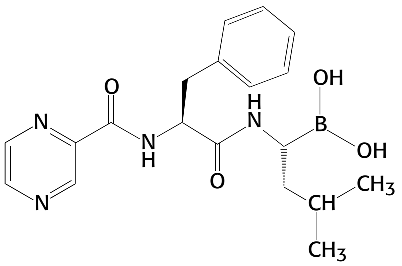
FIGURE 23.8 Proteasome evolution. The archaeal proteasome consists of 14 identical α subunits and 14 identical β subunits. In the eukaryotic proteasome, gene duplication and specialization has led to 7 distinct subunits of each type. The overall architecture of the proteasome is conserved.

Protein degradation can be used to regulate biological function

Table 23.3 lists a number of physiological processes that are controlled at least in part by protein degradation through the ubiquitin–proteasome pathway. In each case, the proteins being degraded are regulatory proteins. Consider, for example, control of the inflammatory response. A transcription factor called *NF-κB* (NF for nuclear factor) initiates the expression of a number of the genes that take part in this response. This factor is itself activated by the degradation of an attached inhibitory protein, *I-κB* (I for inhibitor). In response to inflammatory signals that bind to membrane-bound receptors, I-κB is phosphorylated at two serine residues, creating an E3 binding site. The binding of E3 leads to the ubiquitination and degradation of I-κB, unleashing NF-κB. The liberated transcription factor migrates from the cytoplasm to the nucleus to stimulate the transcription of the target genes. The NF-κB–I-κB system illustrates the interplay of several key regulatory motifs: receptor-mediated signal transduction, phosphorylation, compartmentalization, controlled and specific degradation, and selective gene expression. The importance of the ubiquitin–proteasome system for the regulation of gene expression is highlighted by the use of bortezomib (Velcade), a potent inhibitor of the proteasome, as a therapy for multiple myeloma. Bortezomib is a dipeptidyl boronic acid inhibitor of the proteasome.

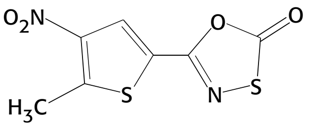
TABLE 23.3 Processes regulated by protein degradation

Gene transcription
Cell-cycle progression
Organ formation
Circadian rhythms
Inflammatory response
Tumor suppression
Cholesterol metabolism
Antigen processing



Bortezomib
(a dipeptidyl boronic acid)

The evolutionary studies of proteasomes described previously have also yielded potential clinical benefits. The bacterial pathogen *Mycobacterium tuberculosis,* the cause of tuberculosis, harbors a proteasome that is very similar to the human counterpart. Nevertheless, recent work has shown that it is possible to exploit the differences between the human and the bacterial proteasomes to develop specific inhibitors of the *M. tuberculosis* complex. Oxathiazol-2-one compounds such as HT1171 are suicide inhibitors (Section 8.5) of the proteolytic activity of the *M. tuberculosis* proteasome, but have no effect on the proteasomes of the human host. This is especially exciting because these drugs kill the nonreplicating form of *M. tuberculosis,* and thus may not require the prolonged treatment required with conventional drugs, thereby reducing the likelihood of drug resistance due to interruption of the treatment regime.



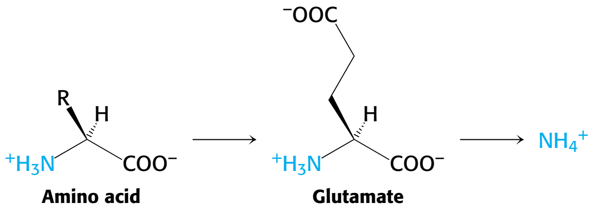
HT1171
[5-(2-methyl-3-nitrothiophen-2-yl)-1,3,4-oxathiazol-2-one]

23.3 The First Step in Amino Acid Degradation Is the Removal of Nitrogen

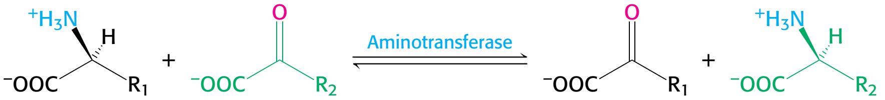
What is the fate of amino acids released on protein digestion or turnover? The first call is for use as building blocks for biosynthetic reactions. However, any not needed as building blocks are degraded to compounds able to enter the metabolic mainstream. The amino group is first removed, and then the remaining carbon skeleton is metabolized to glucose, one of several citric acid cycle intermediates, or to acetyl CoA. The major site of amino acid degradation in mammals is the liver, although muscles readily degrade the branched-chain amino acids (Leu, Ile, and Val). The fate of the α-amino group will be considered first, followed by that of the carbon skeleton (Section 23.5).

Alpha-amino groups are converted into ammonium ions by the oxidative deamination of glutamate

The α-amino group of many amino acids is transferred to α-ketoglutarate to form *glutamate,* which is then oxidatively deaminated to yield ammonium ion (NH_4^+).



Aminotransferases catalyze the transfer of an α-amino group from an α-amino acid to an α-ketoacid. These enzymes, also called *transaminases,* generally funnel α-amino groups from a variety of amino acids to α-ketoglutarate for conversion into NH_4^+.



Aspartate aminotransferase, one of the most important of these enzymes, catalyzes the transfer of the amino group of aspartate to α-ketoglutarate.

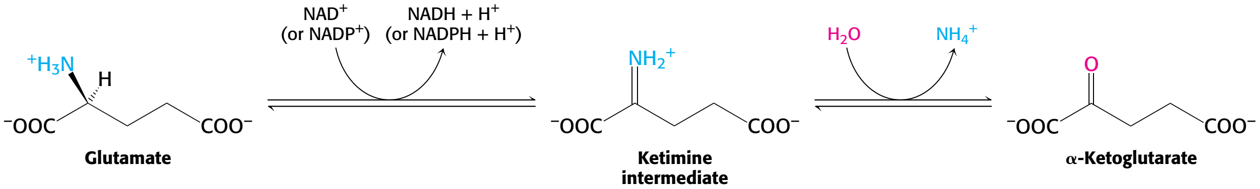
$$\text{Aspartate} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{oxaloacetate} + \text{glutamate}$$

Alanine aminotransferase catalyzes the transfer of the amino group of alanine to α-ketoglutarate.

$$\text{Alanine} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{pyruvate} + \text{glutamate}$$

Transamination reactions are reversible and can thus be used to synthesize amino acids from α-ketoacids, as we shall see in Chapter 24.

The nitrogen atom in glutamate is converted into free ammonium ion by oxidative deamination, a reaction catalyzed by *glutamate dehydrogenase.* This enzyme is unusual in being able to utilize either NAD^+ or $NADP^+$, at least in some species. The reaction proceeds by dehydrogenation of the C—N bond, followed by hydrolysis of the resulting ketimine.



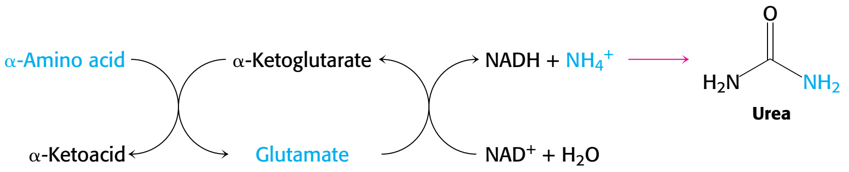
This reaction equilibrium constant is close to 1 in the liver, so the direction of the reaction is determined by the concentrations of reactants and products. Normally, the reaction is driven forward by the rapid removal of ammonium ion. Glutamate dehydrogenase, essentially a liver-specific enzyme, is located in mitochondria, as are some of the other enzymes required for the production of urea. This compartmentalization sequesters free ammonium ion, which is toxic.

In mammals, but not in other organisms, glutamate dehydrogenase is allosterically inhibited by GTP and stimulated by ADP. These nucleotides exert their regulatory effects in a unique manner. An *abortive complex* is formed on the enzyme when a product is replaced by substrate before the reaction is complete. For instance, the enzyme bound to glutamate and NAD(P)H is an abortive complex. GTP facilitates the formation of such complexes while ADP destabilizes the complexes.

The sum of the reactions catalyzed by aminotransferases and glutamate dehydrogenase is

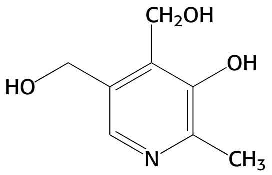
$$\underset{}{\alpha\text{-Amino acid}} + \underset{(\text{or } NADP^+)}{NAD^+} + H_2O \rightleftharpoons \alpha\text{-ketoacid} + NH_4^+ + \underset{(\text{or NADPH})}{NADH} + H^+$$

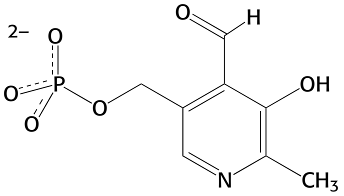
In most terrestrial vertebrates, NH_4^+ is converted into urea, which is excreted.

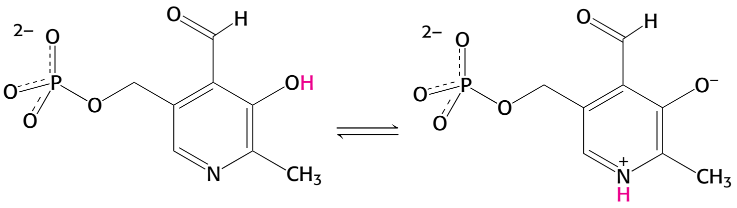


Mechanism: Pyridoxal phosphate forms Schiff-base intermediates in aminotransferases

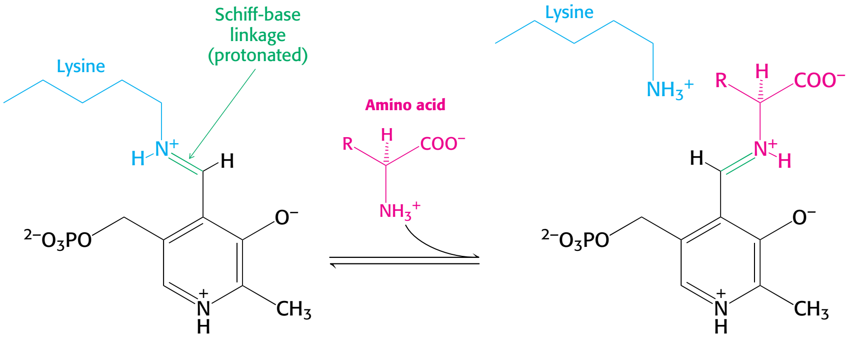
All aminotransferases contain the prosthetic group *pyridoxal phosphate* (PLP), which is derived from *pyridoxine (vitamin B_6)*. Pyridoxal phosphate includes a pyridine ring that is slightly basic to which is attached an OH group that is slightly acidic. Thus, pyridoxal phosphate derivatives can form a stable tautomeric form in which the pyridine nitrogen atom is protonated and, hence, positively charged, while the OH group loses a proton and hence is negatively charged, forming a phenolate.


Pyridoxine (Vitamin B_6)


Pyridoxal phosphate (PLP)


PLP (protonated) **PLP (phenolate)**

The most important functional group on PLP is the aldehyde. This group forms covalent Schiff-base intermediates with amino acid substrates. Indeed, even in the absence of substrate, the aldehyde group of PLP usually forms a Schiff-base linkage with the ε-amino group of a specific lysine residue at the enzyme's active site. A new Schiff-base linkage is formed on addition of an amino acid substrate.



Internal aldimine **External aldimine**

The α-amino group of the amino acid substrate displaces the ε-amino group of the active-site lysine residue. In other words, an *internal* aldimine becomes an *external* aldimine. The amino acid–PLP Schiff base that is formed remains tightly bound to the enzyme by multiple noncovalent interactions. The Schiff-base linkage often accepts a proton at the N, with the positive charge stabilized by interaction with the negatively charged phenolate group of PLP.

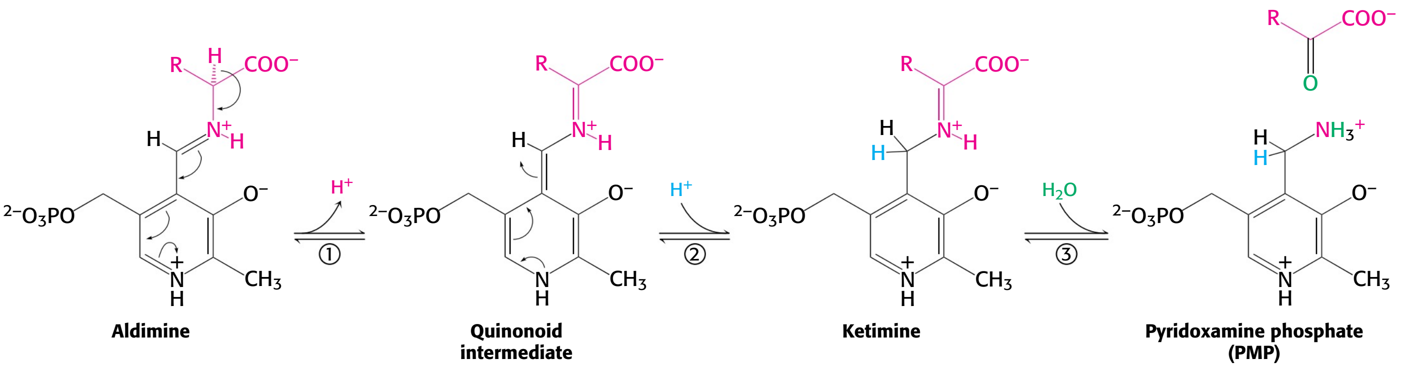
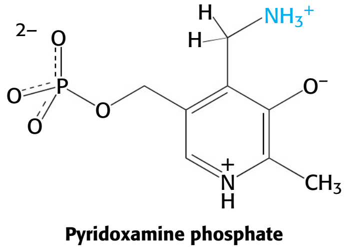


FIGURE 23.9 Transamination mechanism. (1) The external aldimine loses a proton to form a quinonoid intermediate. (2) Reprotonation of this intermediate at the aldehyde carbon atom yields a ketimine. (3) This intermediate is hydrolyzed to generate the α-ketoacid product and pyridoxamine phosphate.

The Schiff base between the amino acid substrate and PLP, the external *aldimine,* loses a proton from the α-carbon atom of the amino acid to form a *quinonoid* intermediate (Figure 23.9). Reprotonation of this intermediate at the aldehyde carbon atom yields a *ketimine.* The ketimine is then hydrolyzed to an α-ketoacid and *pyridoxamine phosphate (PMP).* These steps constitute half of the transamination reaction.

$$\text{Amino acid}_1 + \text{E-PLP} \rightleftharpoons \alpha\text{-ketoacid}_1 + \text{E-PMP}$$

The second half takes place by the reverse of the preceding pathway. A second α-ketoacid reacts with the enzyme–pyridoxamine phosphate complex (E-PMP) to yield a second amino acid and regenerate the enzyme–pyridoxal phosphate complex (E-PLP).



Pyridoxamine phosphate (PMP)

$$\alpha\text{-Ketoacid}_2 + \text{E-PMP} \rightleftharpoons \text{amino acid}_2 + \text{E-PLP}$$

The sum of these partial reactions is

$$\text{Amino acid}_1 + \alpha\text{-ketoacid}_2 \rightleftharpoons \text{amino acid}_2 + \alpha\text{-ketoacid}_1$$

Aspartate aminotransferase is an archetypal pyridoxal-dependent transaminase

The mitochondrial enzyme aspartate aminotransferase provides an especially well studied example of PLP as a coenzyme for transamination reactions (Figure 23.10). X-ray crystallographic studies provided detailed views of how PLP and substrates are bound and confirmed much of the proposed catalytic mechanism. Each of the identical 45-kDa subunits of this dimer

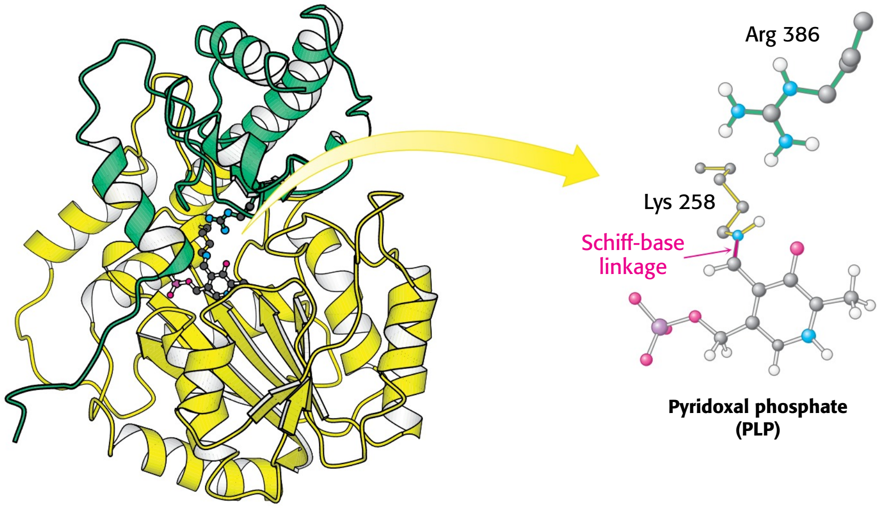


FIGURE 23.10 Aspartate aminotransferase. The active site of this prototypical PLP-dependent enzyme includes pyridoxal phosphate attached to the enzyme by a Schiff-base linkage with lysine 258. An arginine residue in the active site helps orient substrates by binding to their α-carboxylate groups. Only one of the enzyme's two subunits is shown. [Drawn from 1AAW.pdb.]

consists of a large domain and a small one. PLP is bound to the large domain, in a pocket near the domain interface. In the absence of substrate, the aldehyde group of PLP is in a Schiff-base linkage with lysine 258, as expected. Adjacent to the coenzyme's binding site is a conserved arginine residue that interacts with the α-carboxylate group of the amino acid substrate, helping to orient the substrate appropriately in the active site. A base is necessary to remove a proton from the α-carbon group of the amino acid and to transfer it to the aldehyde carbon atom of PLP (Figure 23.9, steps 1 and 2). The lysine amino group that was initially in Schiff-base linkage with PLP appears to serve as the proton donor and acceptor.

Blood levels of aminotransferases serve a diagnostic function

The presence of alanine and aspartate aminotransferase in the blood is an indication of liver damage. Liver damage can occur for a number of reasons, including viral hepatitis, long-term excessive alcohol consumption, and reaction to drugs such as acetaminophen (Section 36.1). Under these conditions, liver cell membranes are damaged and some cellular proteins, including the aminotransferases, leak into the blood. Normal blood values for alanine and aspartate aminotransferase activity are 5–30 units/l and 40–125 units/l, respectively. Depending on the extent of liver damage, the values will reach 200–300 units/l.

Pyridoxal phosphate enzymes catalyze a wide array of reactions

Transamination is just one of a wide range of amino acid transformations that are catalyzed by PLP enzymes. The other reactions catalyzed by PLP enzymes at the α-carbon atom of amino acids are decarboxylations, deaminations, racemizations, and aldol cleavages (Figure 23.11). In addition, PLP enzymes catalyze elimination and replacement reactions at the β-carbon atom (e.g., tryptophan synthetase in the synthesis of tryptophan) and the γ-carbon atom (e.g., cystathionine β-synthase in the synthesis of cysteine) of amino acid substrates. Three common features of PLP catalysis underlie these diverse reactions.

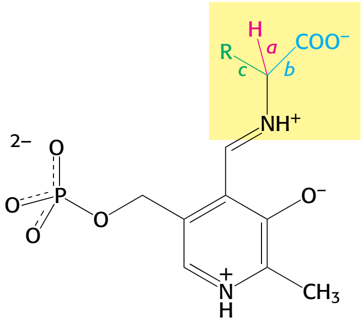


FIGURE 23.11 Bond cleavage by PLP enzymes. Pyridoxal phosphate enzymes labilize one of three bonds at the α-carbon atom of an amino acid substrate. For example, bond *a* is labilized by aminotransferases, bond *b* by decarboxylases, and bond *c* by aldolases (such as threonine aldolases). PLP enzymes also catalyze reactions at the β- and γ-carbon atoms of amino acids.

1. A Schiff base is formed by the amino acid substrate (the amine component) and PLP (the carbonyl component).

2. The protonated form of PLP acts as an *electron sink* to stabilize catalytic intermediates that are negatively charged. Electrons from these intermediates are attracted to the positive charge on the ring nitrogen atom. In other words, PLP is an *electrophilic catalyst.*

3. The product Schiff base is cleaved at the completion of the reaction.

How does an enzyme selectively break a particular one of three bonds at the α-carbon atom of an amino acid substrate? An important principle is that *the bond being broken must be perpendicular to the π orbitals of the electron sink* (Figure 23.12). An aminotransferase, for example, binds the amino acid substrate so that the C_α—H bond is perpendicular to the PLP ring

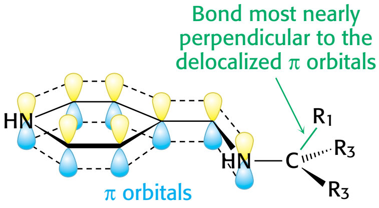

Bond most nearly perpendicular to the delocalized π orbitals
HN
R₁
R₃
HN
C
R₃
π orbitals

FIGURE 23.12 Stereoelectronic effects. The orientation about the N—C_α bond determines the most favored reaction catalyzed by a pyridoxal phosphate enzyme. The bond that is most nearly perpendicular to the plane of delocalized π orbitals (represented by dashed lines) of the pyridoxal phosphate electron sink is most easily cleaved.

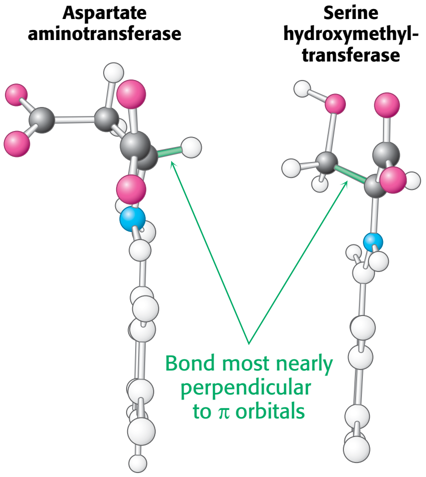

Aspartate aminotransferase
Serine hydroxymethyl-transferase
Bond most nearly perpendicular to π orbitals

FIGURE 23.13 Reaction choice. In aspartate aminotransferase, the C_α—H bond is most nearly perpendicular to the π-orbital system and is cleaved. In serine hydroxymethyltransferase, a small rotation about the N—C_α bond places the C_α—C_β bond perpendicular to the π system, favoring its cleavage.

(Figure 23.13). In serine hydroxymethyltransferase, the enzyme that converts serine into glycine, the N—C_α bond is rotated so that the C_α—C_β bond is most nearly perpendicular to the plane of the PLP ring, favoring its cleavage. This means of choosing one of several possible catalytic outcomes is called *stereoelectronic control.*

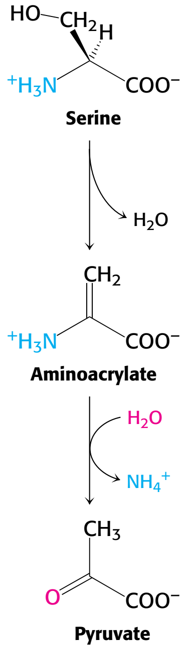
Many of the PLP enzymes that catalyze amino acid transformations, such as serine hydroxymethyltransferase, have a similar structure and are clearly related by divergent evolution. Others, such as tryptophan synthetase, have quite different overall structures. Nonetheless, the active sites of these enzymes are remarkably similar to that of aspartate aminotransferase, revealing the effects of convergent evolution.

Serine and threonine can be directly deaminated

The α-amino groups of serine and threonine can be directly converted into NH_4^+ without first being transferred to α-ketoglutarate. These direct deaminations are catalyzed by *serine dehydratase* and *threonine dehydratase,* in which PLP is the prosthetic group.

$$\text{Serine} \longrightarrow \text{pyruvate} + NH_4^+$$
$$\text{Threonine} \longrightarrow \alpha\text{-ketobutyrate} + NH_4^+$$

These enzymes are called *dehydratases* because *dehydration precedes deamination.* Serine loses a hydrogen ion from its α-carbon atom and a hydroxide ion group from its β-carbon atom to yield aminoacrylate. This unstable compound reacts with H_2O to give pyruvate and NH_4^+. Thus, the presence of a hydroxyl group attached to the β-carbon atom in each of these amino acids permits the direct deamination.



Peripheral tissues transport nitrogen to the liver

Although most amino acid degradation takes place in the liver, other tissues can degrade amino acids. For instance, muscle uses branched-chain amino acids as a source of fuel during prolonged exercise and fasting. How is the nitrogen processed in these other tissues? As in the liver, the first step is the removal of the nitrogen from the amino acid. However, muscle lacks the enzymes of the urea cycle, and so the nitrogen must be released in a nontoxic form that can be absorbed by the liver and converted into urea.

Nitrogen is transported from muscle to the liver in two principal transport forms. Glutamate is formed by transamination reactions, but the nitrogen is then transferred to pyruvate to form alanine, which is released into

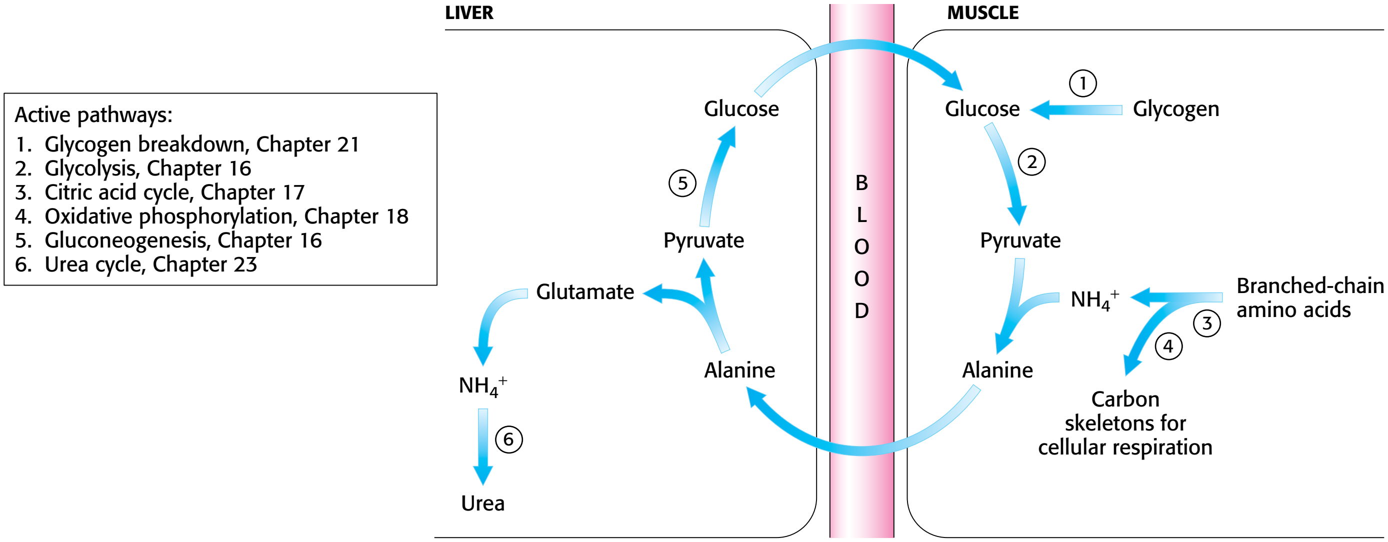

Active pathways:
1. Glycogen breakdown, Chapter 21
2. Glycolysis, Chapter 16
3. Citric acid cycle, Chapter 17
4. Oxidative phosphorylation, Chapter 18
5. Gluconeogenesis, Chapter 16
6. Urea cycle, Chapter 23
LIVER
BLOOD
MUSCLE
Glucose
5
Pyruvate
Glutamate
Alanine
NH4+
6
Urea
Glucose
1
Glycogen
2
Pyruvate
NH4+
Branched-chain amino acids
3
4
Alanine
Carbon skeletons for cellular respiration

FIGURE 23.14 PATHWAY INTEGRATION: The glucose–alanine cycle. During prolonged exercise and fasting, muscle uses branched-chain amino acids as fuel. The nitrogen removed is transferred (through glutamate) to alanine, which is released into the blood stream. In the liver, alanine is taken up and converted into pyruvate for the subsequent synthesis of glucose.

the blood (Figure 23.14). The liver takes up the alanine and converts it back into pyruvate by transamination. The pyruvate can be used for gluconeogenesis and the amino group eventually appears as urea. This transport is referred to as the *glucose–alanine cycle.* It is reminiscent of the Cori cycle discussed earlier (Figure 16.33). However, in contrast with the Cori cycle, pyruvate is not reduced to lactate by NADH, and thus more high-energy electrons are available for oxidative phosphorylation in the muscle.

Glutamine is also a key transport form of nitrogen. Glutamine synthetase catalyzes the synthesis of glutamine from glutamate and NH_4^+ in an ATP-dependent reaction:

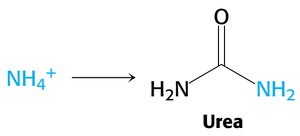
$$NH_4^+ + \text{glutamate} + \text{ATP} \xrightarrow{\text{Glutamine synthetase}} \text{glutamine} + \text{ADP} + P_i$$

The nitrogens of glutamine can be converted into urea in the liver.

23.4 Ammonium Ion Is Converted into Urea in Most Terrestrial Vertebrates

Some of the NH_4^+ formed in the breakdown of amino acids is consumed in the biosynthesis of nitrogen compounds. In most terrestrial vertebrates, the excess NH_4^+ is converted into *urea* and then excreted. Such organisms are referred to as *ureotelic.*

In terrestrial vertebrates, urea is synthesized by the *urea cycle* (Figure 23.15). One of the nitrogen atoms of urea is transferred from an amino acid, aspartate. The other nitrogen atom is derived directly from free NH_4^+, and the carbon atom comes from HCO_3^- (derived by the hydration of CO_2; Section 9.2).



$$NH_4^+ \longrightarrow H_2N{-}C({=}O){-}NH_2$$

Urea

The urea cycle begins with the formation of carbamoyl phosphate

The urea cycle begins with the coupling of free NH_3 with HCO_3^- to form carbamoyl phosphate, the committed reaction of the urea cycle, which is catalyzed by *carbamoyl phosphate*

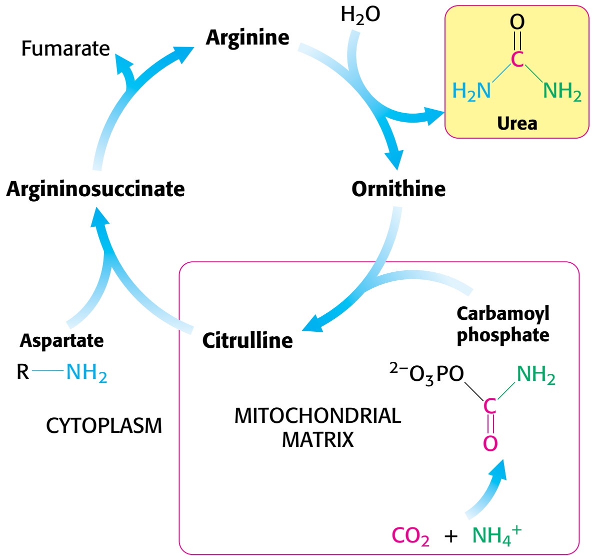
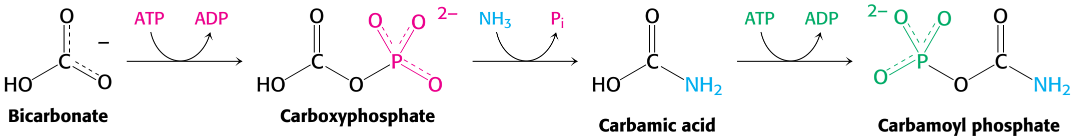

Fumarate
Arginine
H2O
Urea
Argininosuccinate
Ornithine
Aspartate
R—NH2
Citrulline
Carbamoyl phosphate
CYTOPLASM
MITOCHONDRIAL MATRIX
CO2 + NH4+

FIGURE 23.15 The urea cycle.

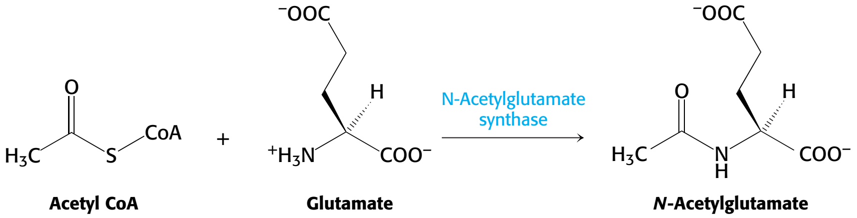
synthetase I. Carbamoyl phosphate is a simple molecule, but its synthesis is complex, requiring three steps.


ATP
ADP
NH_3
P_i
ATP
ADP
Bicarbonate
Carboxyphosphate
Carbamic acid
Carbamoyl phosphate

Note that NH_3, because it is a strong base, normally exists as NH_4^+ in aqueous solution. However, carbamoyl phosphate synthetase uses only NH_3 as a substrate. The reaction begins with the phosphorylation of HCO_3^- to form carboxyphosphate, which then reacts with NH_3 to form carbamic acid. Finally, a second molecule of ATP phosphorylates carbamic acid to form carbamoyl phosphate. The structure and mechanism of the enzyme that catalyzes these reactions will be presented in Chapter 25. The consumption of two molecules of ATP makes this synthesis of carbamoyl phosphate essentially irreversible.

Carbamoyl phosphate synthetase is the key regulatory enzyme for urea synthesis

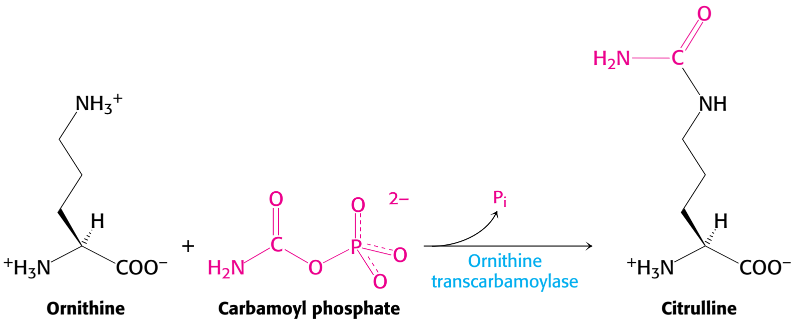
Carbamoyl phosphate synthetase is regulated both allosterically and by covalent modification so that it is maximally active when amino acids are being metabolized for fuel use. The allosteric regulator *N-acetylglutamate* (NAG) is required for synthetase activity. This molecule is synthesized by *N-acetylglutamate synthase*.


N-Acetylglutamate synthase
Acetyl CoA
Glutamate
N-Acetylglutamate

N-acetylglutamate synthase is itself activated by arginine. Thus, NAG is synthesized when amino acids, as represented by arginine and glutamate, are readily available, and carbamoyl phosphate synthetase is then activated to process the generated ammonia. When ammonia is not being generated, the synthetase is inhibited by acetylation. A rise in the mitochondrial NAD^+, indicative of an energy-poor state, stimulates a deacetylase that removes the acetyl group, activating the synthetase and readying the enzyme for processing ammonia from protein degradation. The control of the acetylation of the synthetase is not yet clear.

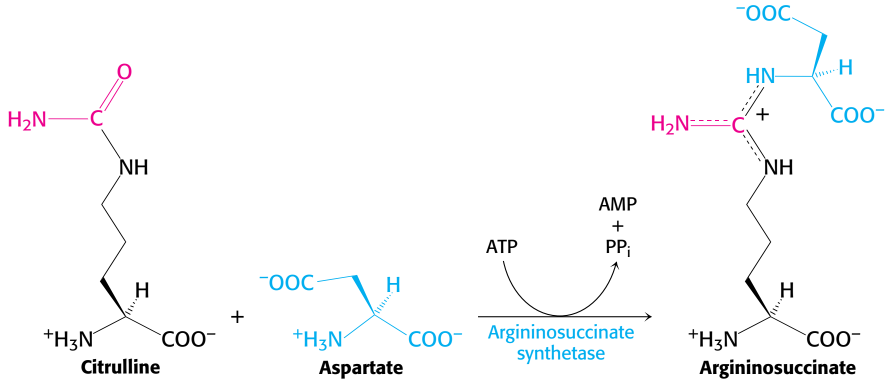
Carbamoyl phosphate reacts with ornithine to begin the urea cycle

The carbamoyl group of carbamoyl phosphate has a high transfer potential because of its anhydride linkage. The carbamoyl group is transferred to *ornithine* to form *citrulline,* in a reaction catalyzed by *ornithine transcarbamoylase.*

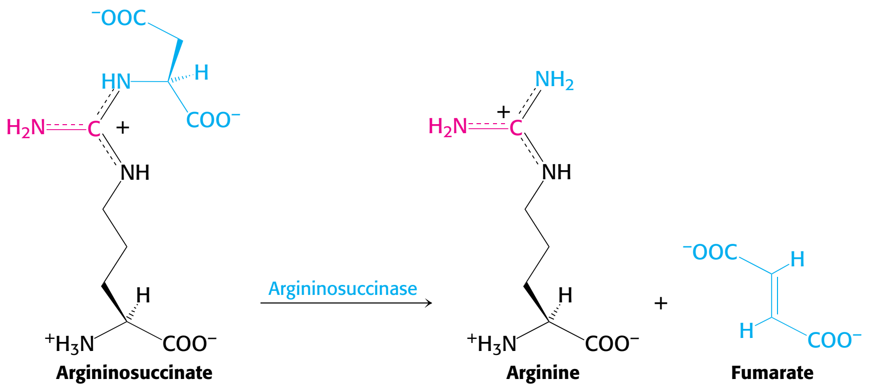

P_i
Ornithine transcarbamoylase
Ornithine
Carbamoyl phosphate
Citrulline

Ornithine and citrulline are amino acids, but they are not used as building blocks of proteins. The formation of NH_4^+ by glutamate dehydrogenase, its incorporation into carbamoyl phosphate as NH_3, and the subsequent synthesis of citrulline take place in the mitochondrial matrix. In contrast, the next three reactions of the urea cycle, which lead to the formation of urea, take place in the cytoplasm.

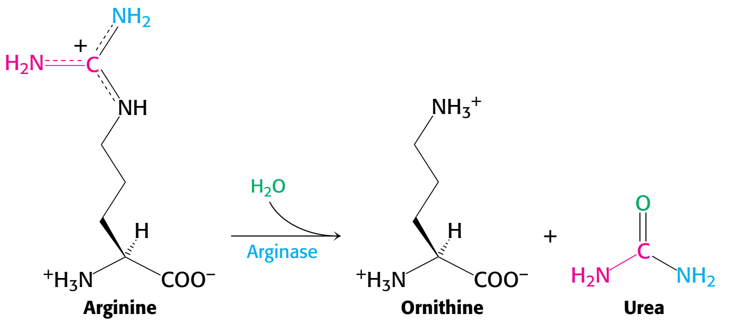
Citrulline is transported to the cytoplasm, where it condenses with aspartate, the donor of the second amino group of urea. This synthesis of *argininosuccinate,* catalyzed by *argininosuccinate synthetase,* is driven by the cleavage of ATP into AMP and pyrophosphate and by the subsequent hydrolysis of pyrophosphate.


Citrulline
Aspartate
ATP
AMP + PPi
Argininosuccinate synthetase
Argininosuccinate

Argininosuccinase (also called argininosuccinate lyase) cleaves argininosuccinate into *arginine* and *fumarate.* Thus, the carbon skeleton of aspartate is preserved in the form of fumarate.


Argininosuccinate
Argininosuccinase
Arginine
Fumarate

Finally, arginine is hydrolyzed to generate urea and ornithine in a reaction catalyzed by *arginase.* Ornithine is then transported back into the mitochondrion to begin another cycle. The urea is excreted. Indeed, human beings excrete about 10 kg (22 pounds) of urea per year.


Arginine
H₂O
Arginase
Ornithine
Urea

In ancient Rome, urine was a valuable commodity. Vessels were placed on street corners for passersby to urinate into. Bacteria would degrade the urea, releasing ammonium ion, which would act as a bleach to brighten togas.

The urea cycle is linked to gluconeogenesis

The stoichiometry of urea synthesis is

$$\mathrm{CO_2} + \mathrm{NH_4}^+ + 3\,\mathrm{ATP} + \text{aspartate} + 2\,\mathrm{H_2O} \longrightarrow \text{urea} + 2\,\mathrm{ADP} + \mathrm{P_i} + \mathrm{AMP} + \mathrm{PP_i} + \text{fumarate}$$

Pyrophosphate is rapidly hydrolyzed, and so the equivalent of four molecules of ATP are consumed in these reactions to synthesize one molecule of urea. The synthesis of fumarate by the urea cycle is important because it is a precursor for glucose synthesis (Figure 23.16). Fumarate is hydrated to malate, which is in turn oxidized to oxaloacetate. Oxaloacetate can be converted into glucose by gluconeogenesis or transaminated to aspartate.

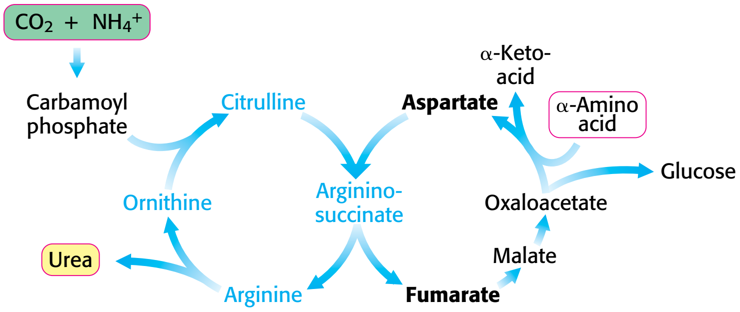


FIGURE 23.16 Metabolic integration of nitrogen metabolism. The urea cycle, gluconeogenesis, and the transamination of oxaloacetate are linked by fumarate and aspartate.

Urea-cycle enzymes are evolutionarily related to enzymes in other metabolic pathways

Carbamoyl phosphate synthetase generates carbamoyl phosphate for both the urea cycle and the first step in pyrimidine biosynthesis (Section 25.1). In mammals, two distinct isozymes of the enzyme are present. Carbamoyl phosphate synthetase II, used in pyrimidine biosynthesis, differs in two important ways from its urea-cycle counterpart. First, this enzyme utilizes glutamine as a nitrogen source rather than $\mathrm{NH_3}$. The side-chain amide of glutamine is hydrolyzed within one domain of the enzyme, and the ammonia generated moves through a tunnel in the enzyme to a second active site, where it reacts with carboxyphosphate. Second, this enzyme is part of a large complex that catalyzes several steps in pyrimidine biosynthesis (Section 25.1). Interestingly, the domain in which glutamine hydrolysis takes place is largely preserved in the urea-cycle enzyme, although that domain is catalytically inactive. This site binds *N*-acetylglutamate, an allosteric activator of the enzyme. *A catalytic site in one isozyme has been adapted to act as an allosteric site in another isozyme having a different physiological role.*

Can we find homologs for the other enzymes in the urea cycle? Ornithine transcarbamoylase is homologous to aspartate transcarbamoylase, which catalyzes the first step in pyrimidine biosynthesis, and the structures of their catalytic subunits are quite similar (Figure 23.17). Thus, two consecutive steps in the pyrimidine biosynthetic pathway were adapted for urea synthesis. The next step in the urea cycle is the addition of aspartate to citrulline to form argininosuccinate, and the subsequent step is the removal of fumarate. These two steps together accomplish the net addition of an amino group to citrulline to form arginine. Remarkably, these steps are analogous to two consecutive steps in the purine biosynthetic pathway (Section 25.2).

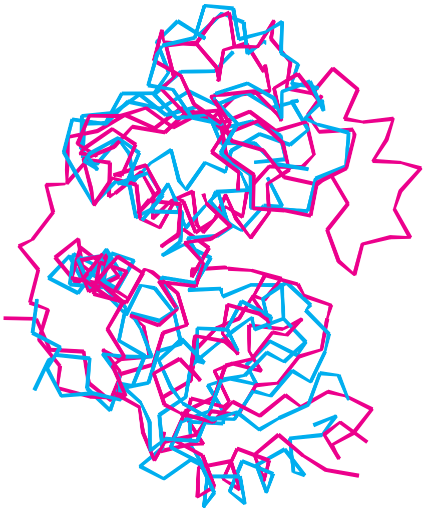
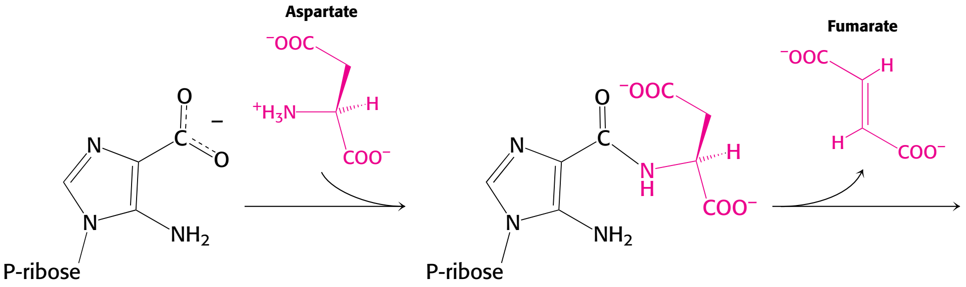


FIGURE 23.17 Homologous enzymes. The structure of the catalytic subunit of ornithine transcarbamoylase (blue) is quite similar to that of the catalytic subunit of aspartate transcarbamoylase (red), indicating that these two enzymes are homologs. [Drawn from 1AKM.pdb and 1RAI.pdb.]



The enzymes that catalyze these steps are homologous to argininosuccinate synthetase and argininosuccinase, respectively. Thus, four of the five enzymes in the urea cycle were adapted from enzymes taking part in nucleotide biosynthesis. The remaining enzyme, arginase, appears to be an ancient enzyme found in all domains of life.

Inherited defects of the urea cycle cause hyperammonemia and can lead to brain damage

The synthesis of urea in the liver is the major route for the removal of NH_4^+. Urea cycle disorders occur with a prevalence of about 1 in 15,000. A blockage of carbamoyl phosphate synthesis or of any of the four steps of the urea cycle has devastating consequences because there is no alternative pathway for the synthesis of urea. *All defects in the urea cycle lead to an elevated level of* NH_4^+ *in the blood (hyperammonemia).* Some of these genetic defects become evident a day or two after birth, when the afflicted infant becomes lethargic and vomits periodically. Coma and irreversible brain damage may soon follow, a condition called *hepatic encephalopathy*. Why are high levels of NH_4^+ toxic? The answer to this question is not yet known. Recent work, however, suggests that NH_4^+ may inappropriately activate a sodium-potassium-chloride cotransporter. This activation disrupts the osmotic balance of the nerve cell, causing swelling that damages the cell and results in neurological disorders.

Ingenious strategies for coping with deficiencies in urea synthesis have been devised on the basis of a thorough understanding of the underlying biochemistry. Consider, for example, *argininosuccinase deficiency.* This defect can be partly bypassed by *providing a surplus of arginine in the diet and restricting the total protein intake.* In the liver, arginine is split into urea and ornithine, which then reacts with carbamoyl phosphate to form citrulline (Figure 23.18). This urea-cycle intermediate condenses with aspartate to yield argininosuccinate, which is then excreted. Note that two nitrogen atoms—one from carbamoyl phosphate and the other from aspartate—are eliminated from the body per molecule of arginine provided in the diet. In essence, *argininosuccinate substitutes for urea in carrying nitrogen out of the body.*

The treatment of *carbamoyl phosphate synthetase deficiency* or *ornithine transcarbamoylase deficiency* illustrates a different strategy for circumventing a metabolic block. Citrulline and argininosuccinate cannot be used to dispose of nitrogen atoms because their formation is impaired. Under these conditions, excess nitrogen accumulates in glycine and glutamine. The challenge then is to rid the body of the nitrogen accumulating in these two amino acids. That goal is accomplished by supplementing a protein-restricted diet with *large amounts of benzoate and phenylacetate.* Benzoate is activated to benzoyl CoA, which reacts with glycine to form hippurate, which is excreted (Figure 23.19). Likewise, phenylacetate is activated to phenylacetyl CoA, which reacts with glutamine to form phenylacetylglutamine, which is also excreted. These conjugates substitute for urea in the

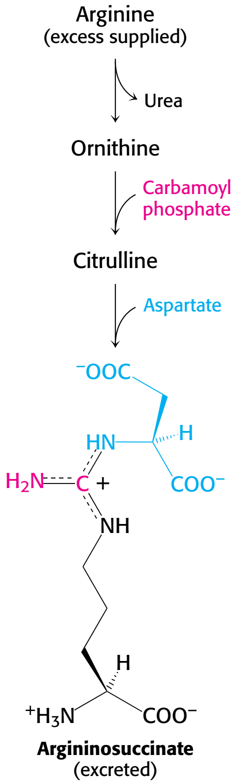


FIGURE 23.18 Treatment of argininosuccinase deficiency. Argininosuccinase deficiency can be managed by supplementing the diet with arginine. Nitrogen is excreted in the form of argininosuccinate.

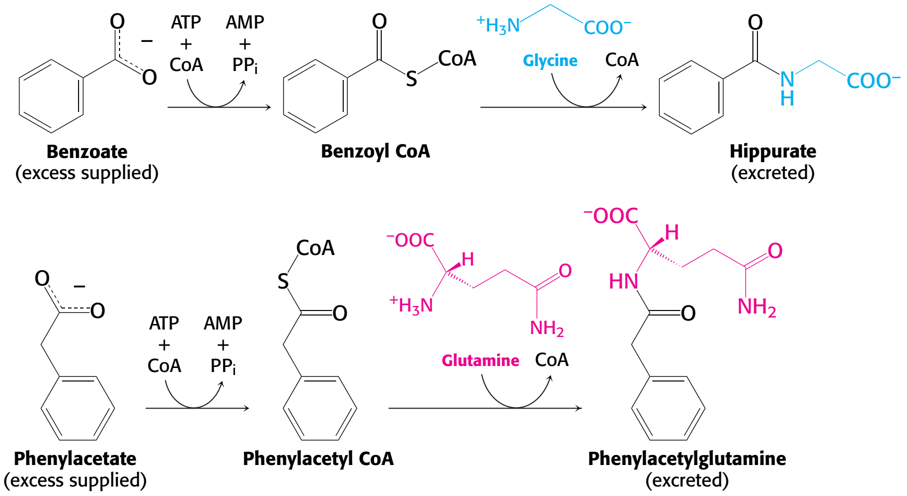


FIGURE 23.19 Treatment of carbamoyl phosphate synthetase and ornithine transcarbamoylase deficiencies. Both deficiencies can be treated by supplementing the diet with benzoate and phenylacetate. Nitrogen is excreted in the form of hippurate and phenylacetylglutamine.

disposal of nitrogen. Thus, *latent biochemical pathways can be activated to partly bypass a genetic defect.*

Urea is not the only means of disposing of excess nitrogen

As stated earlier, most terrestrial vertebrates are ureotelic; they excrete excess nitrogen as urea. However, urea is not the only excretable form of nitrogen. *Ammoniotelic organisms, such as aquatic vertebrates and invertebrates, release nitrogen as* NH_4^+ and rely on the aqueous environment to dilute this toxic substance. Interestingly, lungfish, which are normally ammoniotelic, become ureotelic in time of drought, when they live out of the water.

Both ureotelic and ammoniotelic organisms depend on sufficient water, to varying degrees, for nitrogen excretion. *In contrast, uricotelic organisms, such as birds and reptiles, secrete nitrogen as the purine uric acid.* Uric acid is secreted as an almost solid slurry requiring little water. The secretion of uric acid also has the advantage of removing four atoms of nitrogen per molecule. The pathway for nitrogen excretion developed in the course of evolution clearly depends on the habitat of the organism.

23.5 Carbon Atoms of Degraded Amino Acids Emerge as Major Metabolic Intermediates

We now turn to the fates of the carbon skeletons of amino acids after the removal of the α-amino group. *The strategy of amino acid degradation is to transform the carbon skeletons into major metabolic intermediates that can be converted into glucose or oxidized by the citric acid cycle.* The conversion pathways range from extremely simple to quite complex. The carbon skeletons of the diverse set of 20 fundamental amino acids are funneled into only seven molecules: *pyruvate, acetyl CoA, acetoacetyl CoA, α-ketoglutarate, succinyl CoA, fumarate,* and *oxaloacetate.* We see here an example of the remarkable economy of metabolic conversions.

Amino acids that are degraded to acetyl CoA or acetoacetyl CoA are termed *ketogenic* amino acids because they can give rise to ketone bodies or fatty acids. Amino acids that are degraded to pyruvate, α-ketoglutarate, succinyl CoA, fumarate, or oxaloacetate are termed *glucogenic* amino acids. Oxaloacetate, generated from pyruvate and other citric acid cycle intermediates, can be converted into phosphoenolpyruvate and then into glucose (Section 16.3). Recall that mammals lack a pathway for the net synthesis of glucose from acetyl CoA or acetoacetyl CoA.

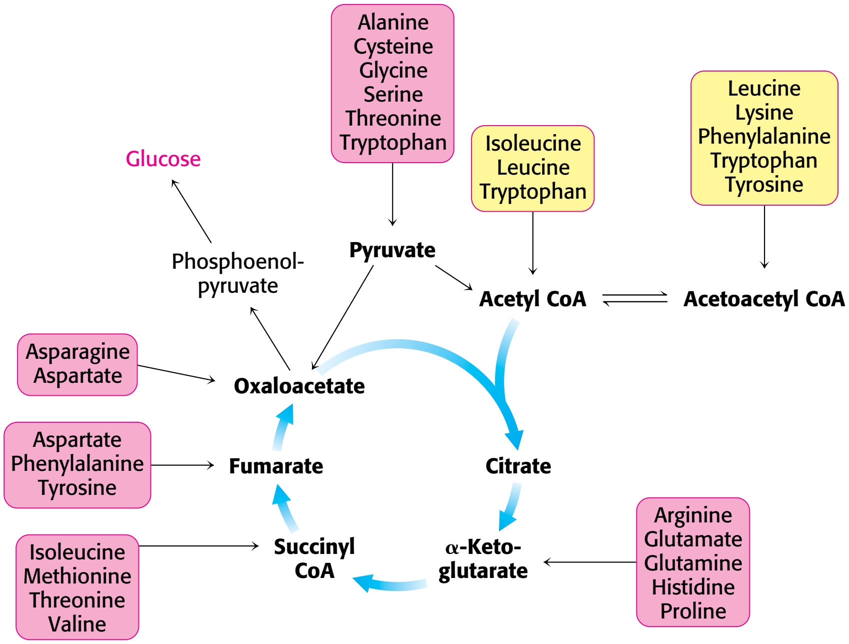

Alanine
Cysteine
Glycine
Serine
Threonine
Tryptophan
Isoleucine
Leucine
Tryptophan
Leucine
Lysine
Phenylalanine
Tryptophan
Tyrosine
Glucose
Phosphoenol-pyruvate
Pyruvate
Acetyl CoA
Acetoacetyl CoA
Asparagine
Aspartate
Oxaloacetate
Aspartate
Phenylalanine
Tyrosine
Fumarate
Citrate
Isoleucine
Methionine
Threonine
Valine
Succinyl CoA
α-Ketoglutarate
Arginine
Glutamate
Glutamine
Histidine
Proline

FIGURE 23.20 Fates of the carbon skeletons of amino acids. Glucogenic amino acids are shaded red, and ketogenic amino acids are shaded yellow. Several amino acids are both glucogenic and ketogenic.

Of the basic set of 20 amino acids, only leucine and lysine are solely ketogenic (Figure 23.20). Isoleucine, phenylalanine, tryptophan, and tyrosine are both ketogenic and glucogenic. Some of their carbon atoms emerge in acetyl CoA or acetoacetyl CoA, whereas others appear in potential precursors of glucose. The other 14 amino acids are classed as solely glucogenic. We will identify the degradation pathways by the entry point into metabolism.

Pyruvate is an entry point into metabolism for a number of amino acids

Pyruvate is the entry point of the three-carbon amino acids—alanine, serine, and cysteine—into the metabolic mainstream (Figure 23.21). The transamination of alanine directly yields pyruvate.

$$\text{Alanine} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{pyruvate} + \text{glutamate}$$

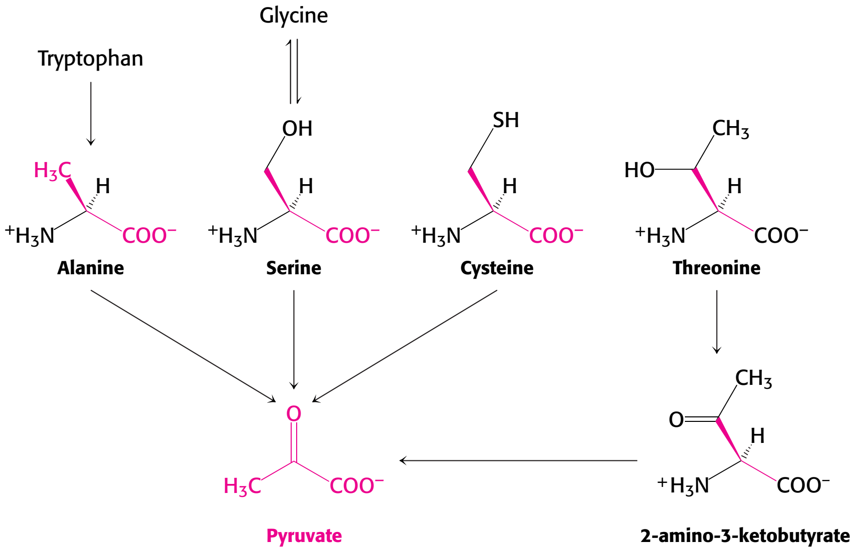

Glycine
Tryptophan
Alanine
Serine
Cysteine
Threonine
Pyruvate
2-amino-3-ketobutyrate

FIGURE 23.21 Pyruvate formation from amino acids. Pyruvate is the point of entry for alanine, serine, cysteine, glycine, threonine, and tryptophan.

As mentioned earlier in the chapter, glutamate is then oxidatively deaminated, yielding NH_4^+ and regenerating α-ketoglutarate. The sum of these reactions is

$$\text{Alanine} + \text{NAD(P)}^+ + H_2O \rightleftharpoons \text{pyruvate} + NH_4^+ + \text{NAD(P)H} + H^+$$

Another simple reaction in the degradation of amino acids is the *deamination of serine to pyruvate* by *serine dehydratase* (p. 692).

$$\text{Serine} \longrightarrow \text{pyruvate} + NH_4^+$$

Cysteine can be converted into pyruvate by several pathways, with its sulfur atom emerging in H_2S, SCN^-, or SO_3^{2-}.

The carbon atoms of three other amino acids can be converted into pyruvate. *Glycine* can be converted into serine by the enzymatic addition of a hydroxymethyl group or it can be cleaved to give CO_2, NH_4^+, and an activated one-carbon unit. *Threonine* can give rise to pyruvate through the intermediate 2-amino-3-ketobutyrate. Three carbon atoms of *tryptophan* can emerge in alanine, which can be converted into pyruvate.

Oxaloacetate is an entry point into metabolism for aspartate and asparagine

Aspartate and asparagine are converted into oxaloacetate, a citric acid cycle intermediate. *Aspartate,* a four-carbon amino acid, is directly *transaminated to oxaloacetate.*

$$\text{Aspartate} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{oxaloacetate} + \text{glutamate}$$

Asparagine is hydrolyzed by *asparaginase* to NH_4^+ and aspartate, which is then transaminated.

Recall that aspartate can also be converted into *fumarate* by the urea cycle (Figure 23.16). Fumarate is a point of entry for half the carbon atoms of tyrosine and phenylalanine, as will be discussed shortly.

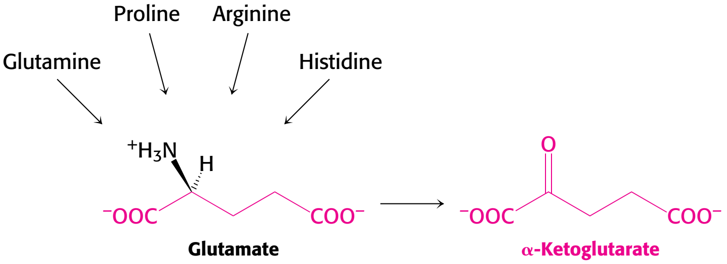


FIGURE 23.22 α-Ketoglutarate formation from amino acids. α-Ketoglutarate is the point of entry of several five-carbon amino acids that are first converted into glutamate.

Alpha-ketoglutarate is an entry point into metabolism for five-carbon amino acids

The carbon skeletons of several five-carbon amino acids enter the citric acid cycle at *α-ketoglutarate.* These amino acids are first converted into *glutamate,* which is then oxidatively deaminated by glutamate dehydrogenase to yield α-ketoglutarate (Figure 23.22).

Histidine is converted into 4-imidazolone 5-propionate (Figure 23.23). The amide bond in the ring of this intermediate is hydrolyzed to the *N*-formimino derivative of glutamate, which is then converted into glutamate by the transfer of its formimino group to tetrahydrofolate, a carrier of activated one-carbon units (Figure 24.9).

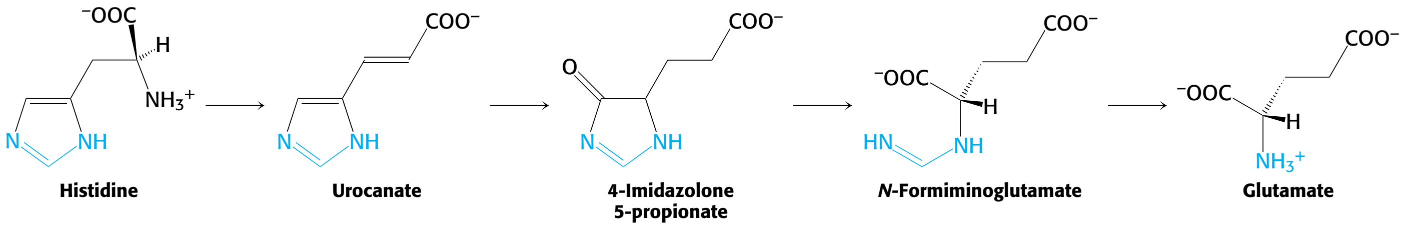


FIGURE 23.23 Histidine degradation. Conversion of histidine into glutamate.

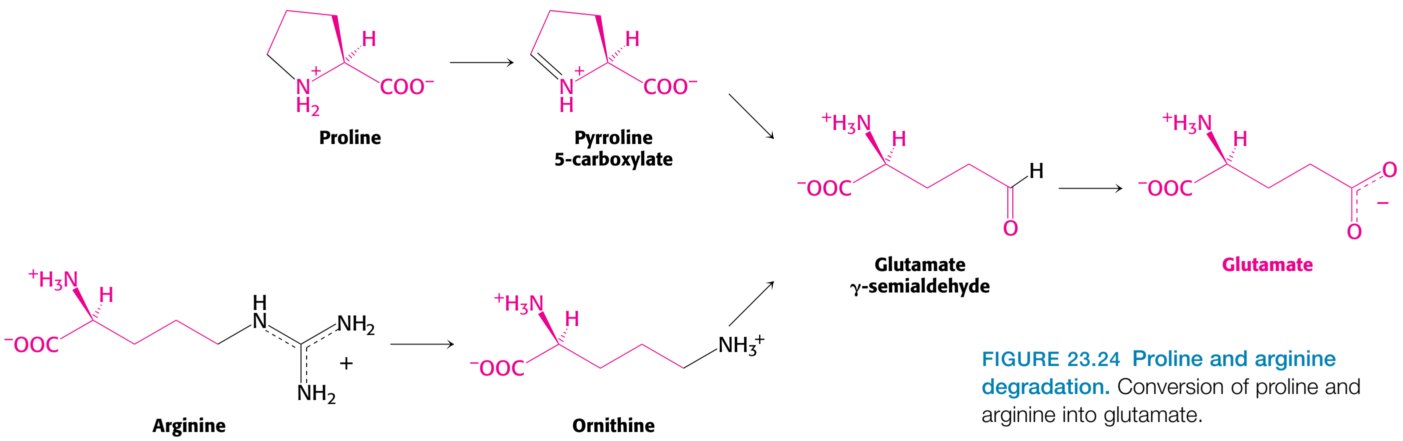


FIGURE 23.24 Proline and arginine degradation. Conversion of proline and arginine into glutamate.

Glutamine is hydrolyzed to glutamate and NH_4^+ by *glutaminase. Proline* and *arginine* are each converted into glutamate γ-semialdehyde, which is then oxidized to glutamate (Figure 23.24).

Succinyl coenzyme A is a point of entry for several nonpolar amino acids

Succinyl CoA is a point of entry for some of the carbon atoms of methionine, isoleucine, and valine. Propionyl CoA and then methylmalonyl CoA are intermediates in the breakdown of these three nonpolar amino acids (Figure 23.25). This pathway from propionyl CoA to succinyl CoA is also used in the oxidation of fatty acids that have an odd number of carbon atoms. The mechanism for the interconversion of propionyl CoA and methylmalonyl CoA was presented in Section 22.3.

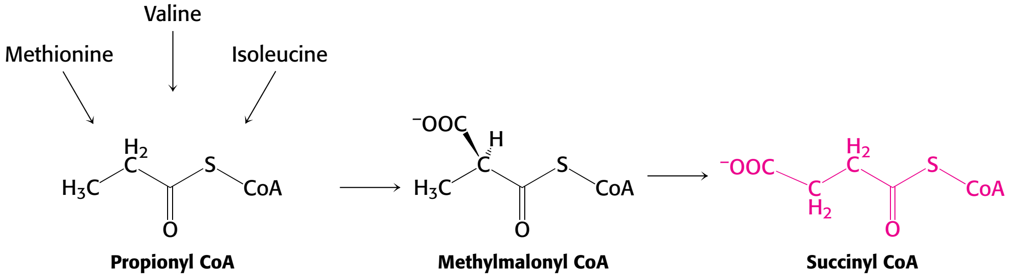


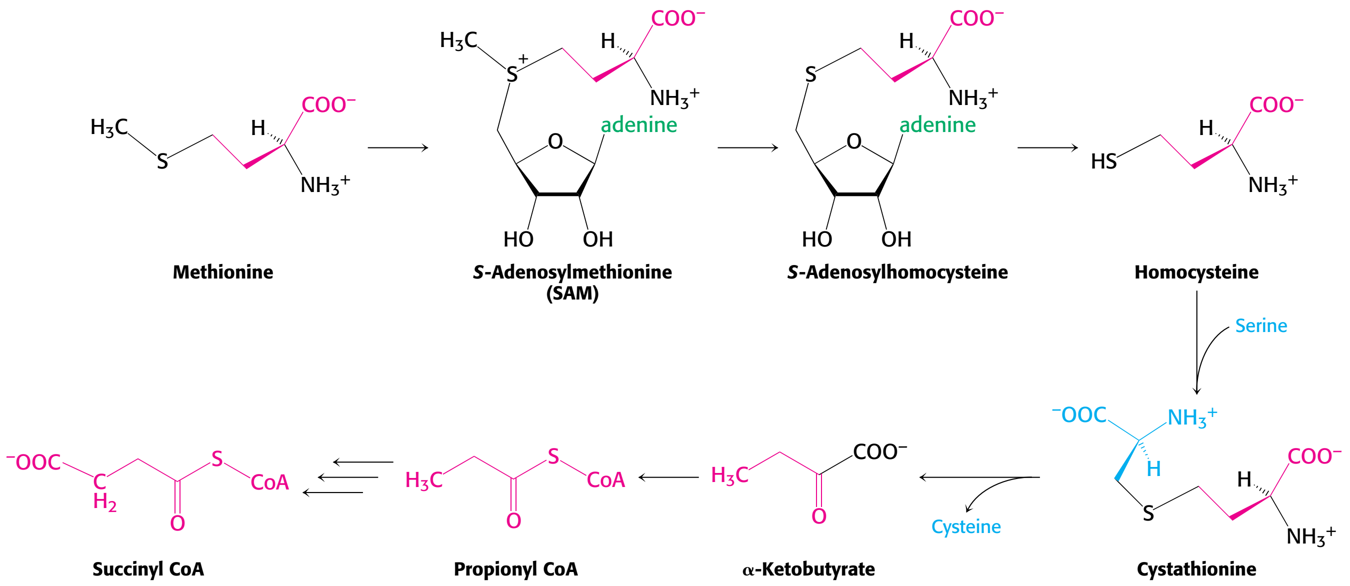
FIGURE 23.25 Succinyl CoA formation. Conversion of methionine, isoleucine, and valine into succinyl CoA.

Methionine degradation requires the formation of a key methyl donor, *S*-adenosylmethionine

Methionine is converted into succinyl CoA in nine steps (Figure 23.26). The first step is the adenylation of methionine to form S-*adenosylmethionine* (SAM), a common methyl donor in the cell (Section 24.2). Loss of the methyl and adenosyl groups yields homocysteine, which is eventually processed to *α-ketobutyrate.* This α-ketoacid is oxidatively decarboxylated by the α-ketoacid dehydrogenase complex to *propionyl CoA*, which is processed to *succinyl CoA,* as described in Section 22.3.

The branched-chain amino acids yield acetyl CoA, acetoacetate, or propionyl CoA

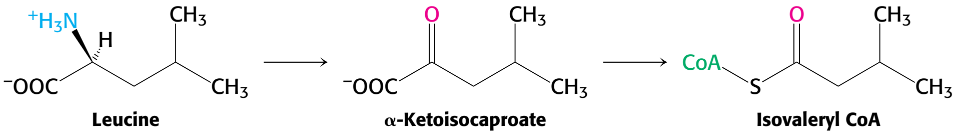
The branched-chain amino acids are degraded by reactions that we have already encountered in the citric acid cycle and fatty acid oxidation. Leucine is transaminated to the corresponding α-ketoacid, *α-ketoisocaproate.*



Methionine
S-Adenosylmethionine (SAM)
adenine
S-Adenosylhomocysteine
Homocysteine
Serine
Cystathionine
Cysteine
α-Ketobutyrate
Propionyl CoA
Succinyl CoA

FIGURE 23.26 Methionine metabolism. The pathway for the conversion of methionine into succinyl CoA. *S*-Adenosylmethionine, formed along this pathway, is an important methyl donor.

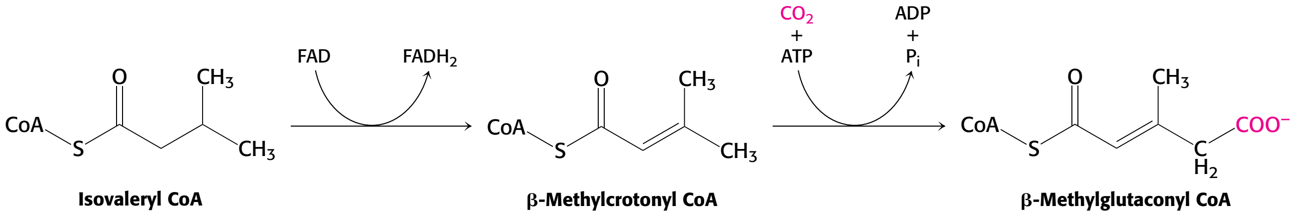
This α-ketoacid is *oxidatively decarboxylated to isovaleryl CoA* by the *branched-chain α-ketoacid dehydrogenase complex.*



Leucine
α-Ketoisocaproate
Isovaleryl CoA

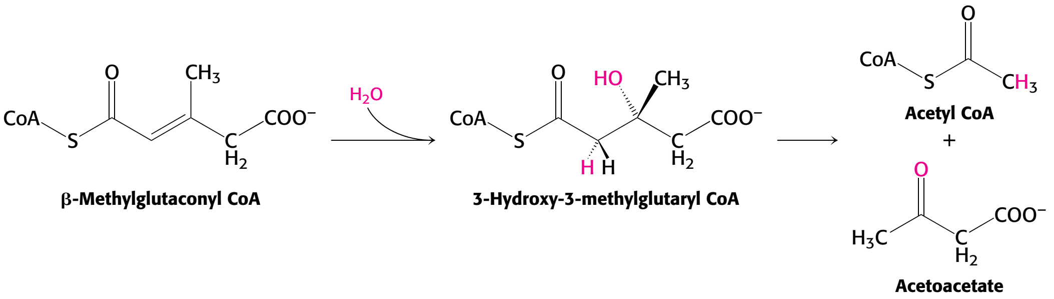
The α-ketoacids of valine and isoleucine, the other two branched-chain aliphatic amino acids, also are substrates (as is α-ketobutyrate derived from methionine). The oxidative decarboxylation of these α-ketoacids is analogous to that of pyruvate to acetyl CoA and of α-ketoglutarate to succinyl CoA. The branched-chain α-ketoacid dehydrogenase, a multienzyme complex, is a homolog of pyruvate dehydrogenase (Section 17.1) and α-ketoglutarate dehydrogenase (Section 17.2). Indeed, the E3 components of these enzymes, which regenerate the oxidized form of lipoamide, are identical.

The isovaleryl CoA derived from leucine is *dehydrogenated* to yield β-*methylcrotonyl CoA.* This oxidation is catalyzed by *isovaleryl CoA dehydrogenase.* The hydrogen acceptor is FAD, as in the analogous reaction in fatty acid oxidation that is catalyzed by acyl CoA dehydrogenase. *β-Methylglutaconyl CoA* is then formed by the *carboxylation* of β-methylcrotonyl CoA at the expense of the hydrolysis of a molecule of ATP. As might be expected, the carboxylation mechanism of β-methylcrotonyl CoA carboxylase is similar to that of pyruvate carboxylase and acetyl CoA carboxylase.



FAD
FADH2
CO2 + ATP
ADP + Pi
Isovaleryl CoA
β-Methylcrotonyl CoA
β-Methylglutaconyl CoA

β-Methylglutaconyl CoA is then *hydrated* to form *3-hydroxy-3-methylglutaryl CoA,* which is cleaved into *acetyl CoA* and *acetoacetate.* This reaction has already been discussed in regard to the formation of ketone bodies from fatty acids (Section 22.3).

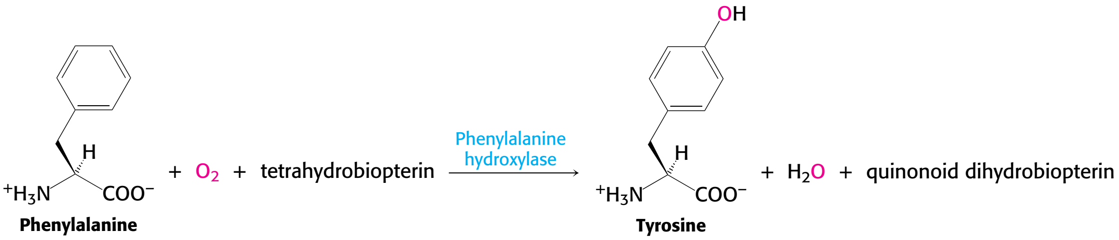

β-Methylglutaconyl CoA
H2O
3-Hydroxy-3-methylglutaryl CoA
Acetyl CoA
+
Acetoacetate

The degradative pathways of valine and isoleucine resemble that of leucine. After transamination and oxidative decarboxylation to yield a CoA derivative, the subsequent reactions are like those of fatty acid oxidation. Isoleucine yields acetyl CoA and propionyl CoA, whereas valine yields CO_2 and propionyl CoA. The degradation of leucine, valine, and isoleucine validate a point made earlier (Chapter 15): the number of reactions in metabolism is large, but the number of *kinds* of reactions is relatively small. The degradation of leucine, valine, and isoleucine provides a striking illustration of the underlying simplicity and elegance of metabolism.

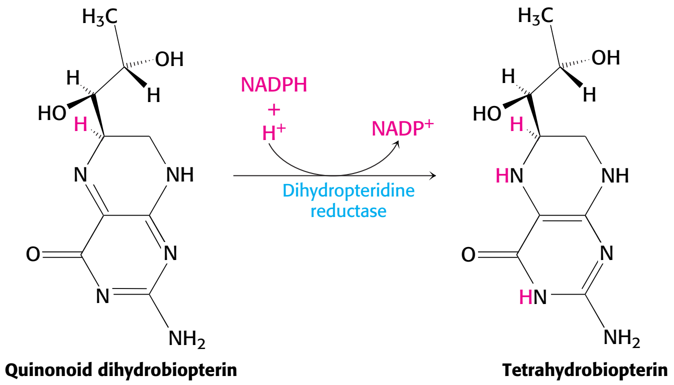
Oxygenases are required for the degradation of aromatic amino acids

The degradation of the aromatic amino acids yields the common intermediates acetoacetate, fumarate, and pyruvate. The degradation pathway is not as straightforward as that of the amino acids previously discussed. For the aromatic amino acids, *molecular oxygen is used to break the aromatic ring.*

The degradation of phenylalanine begins with its hydroxylation to tyrosine, a reaction catalyzed by *phenylalanine hydroxylase.* This enzyme is called a *monooxygenase* (or *mixed-function oxygenase*) because *one atom of* O_2 appears in the product and the other in H_2O.


Phenylalanine
+ O2 + tetrahydrobiopterin
Phenylalanine hydroxylase
Tyrosine
+ H2O + quinonoid dihydrobiopterin

The reductant here is *tetrahydrobiopterin,* an electron carrier that has not been previously discussed and is derived from the cofactor *biopterin*. Because biopterin is synthesized in the body, it is not a vitamin. The quinonoid form of dihydrobiopterin is produced in the hydroxylation of phenylalanine. It is reduced back to tetrahydrobiopterin by NADPH in a reaction catalyzed by *dihydropteridine reductase*.


NADPH
+
H+
NADP+
Dihydropteridine reductase
Quinonoid dihydrobiopterin
Tetrahydrobiopterin

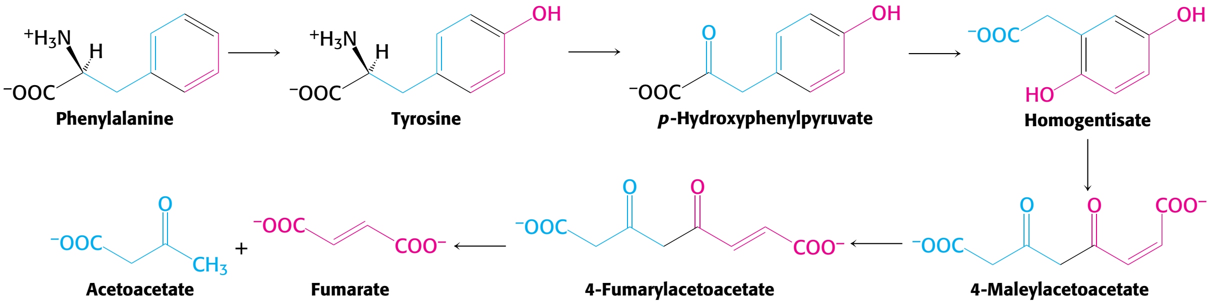

Phenylalanine
Tyrosine
p-Hydroxyphenylpyruvate
Homogentisate
4-Maleylacetoacetate
4-Fumarylacetoacetate
Fumarate
Acetoacetate

FIGURE 23.27 Phenylalanine and tyrosine degradation. The pathway for the conversion of phenylalanine into acetoacetate and fumarate.

The sum of the reactions catalyzed by phenylalanine hydroxylase and dihydropteridine reductase is

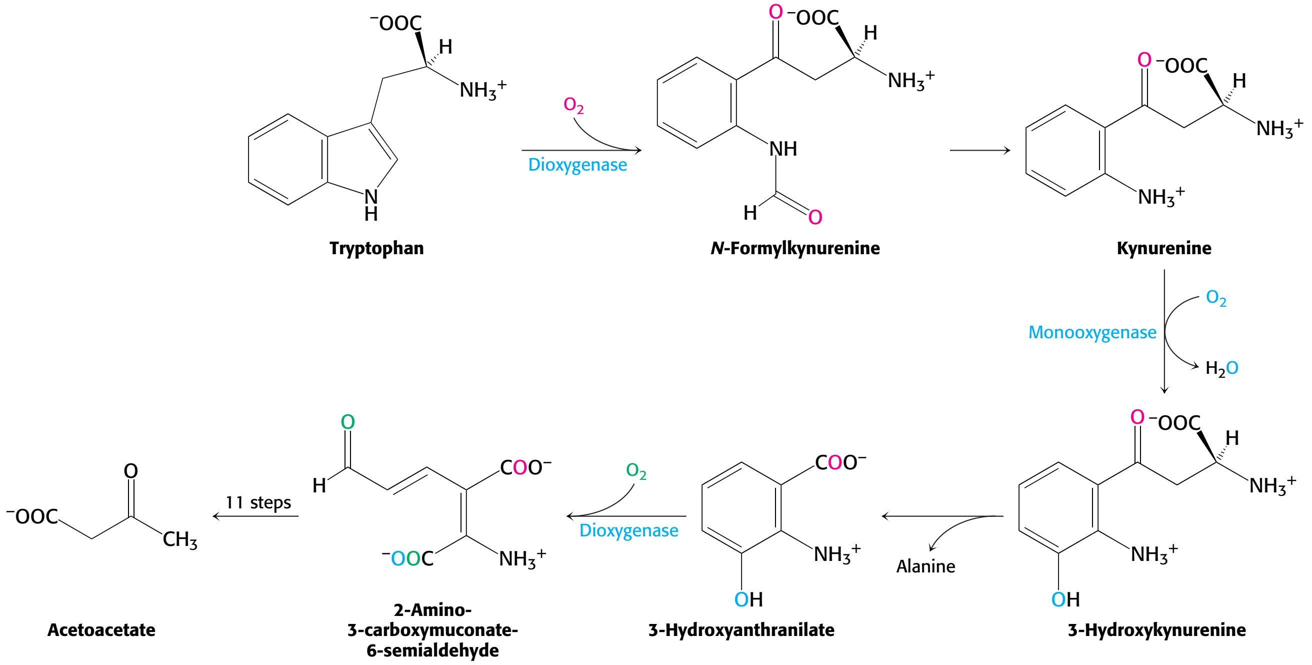
$$\text{Phenylalanine} + \mathrm{O_2} + \mathrm{NADPH} + \mathrm{H^+} \longrightarrow \text{tyrosine} + \mathrm{NADP^+} + \mathrm{H_2O}$$

Note that these reactions can also be used to synthesize tyrosine from phenylalanine.

The next step in the degradation of phenylalanine and tyrosine is the transamination of tyrosine to p-*hydroxyphenylpyruvate* (Figure 23.27). This α-ketoacid then reacts with O_2 to form *homogentisate.* The enzyme catalyzing this complex reaction, p-*hydroxyphenylpyruvate hydroxylase,* is called a *dioxygenase* because *both atoms of* O_2 become incorporated into the product, one on the ring and one in the carboxyl group. The aromatic ring of homogentisate is then cleaved by O_2, which yields 4-maleylacetoacetate. This reaction is catalyzed by *homogentisate oxidase,* another dioxygenase. 4-Maleylacetoacetate is then isomerized to *4-fumarylacetoacetate* by an enzyme that uses glutathione as a cofactor. Finally, 4-fumarylacetoacetate is hydrolyzed to *fumarate* and *acetoacetate.*

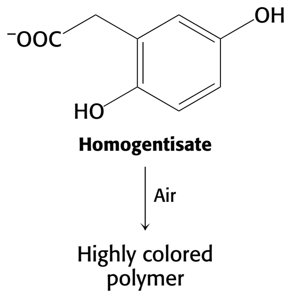
Tryptophan degradation requires several oxygenases (Figure 23.28). Tryptophan 2,3-dioxygenase cleaves the pyrrole ring, and kynureinine 3-monooxygenase hydroxylates the remaining benzene ring, a reaction similar to the hydroxylation of phenylalanine to form tyrosine. Alanine is removed and the 3-hydroxyanthranilate is cleaved by another dioxygenase and

FIGURE 23.28 Tryptophan degradation. The pathway for the conversion of tryptophan into alanine and acetoacetate.


Tryptophan
O2
Dioxygenase
N-Formylkynurenine
Kynurenine
O2
Monooxygenase
H2O
3-Hydroxykynurenine
Alanine
3-Hydroxyanthranilate
O2
Dioxygenase
2-Amino-3-carboxymuconate-6-semialdehyde
11 steps
Acetoacetate

subsequently processed to acetoacetyl CoA. *Nearly all cleavages of aromatic rings in biological systems are catalyzed by dioxygenases*. The active sites of these enzymes contain iron that is not part of heme or an iron–sulfur cluster.

23.6 Inborn Errors of Metabolism Can Disrupt Amino Acid Degradation



Errors in amino acid metabolism provided some of the first examples of biochemical defects linked to pathological conditions. For instance, *alcaptonuria,* an inherited metabolic disorder caused by the absence of homogentisate oxidase, was described in 1902. Homogentisate, a normal intermediate in the degradation of phenylalanine and tyrosine (Figure 23.27), accumulates in alcaptonuria because its degradation is blocked. Homogentisate is excreted in the urine, which turns dark on standing as homogentisate is oxidized and polymerized to a melanin-like substance.

Although alcaptonuria is a relatively harmless condition, such is not the case with other errors in amino acid metabolism. In *maple syrup urine disease,* the oxidative decarboxylation of α-ketoacids derived from valine, isoleucine, and leucine is blocked because the branched-chain dehydrogenase is missing or defective. Hence, the levels of these α-ketoacids and the branched-chain amino acids that give rise to them are markedly elevated in both blood and urine. The urine of patients has the odor of maple syrup—hence the name of the disease (also called *branched-chain ketoaciduria*). Maple syrup urine disease usually leads to mental and physical retardation unless the patient is placed on a diet low in valine, isoleucine, and leucine early in life. The disease can be readily detected in newborns by screening urine samples with 2,4-dinitrophenylhydrazine, which reacts with α-ketoacids to form 2,4-dinitrophenylhydrazone derivatives. A definitive diagnosis can be made by mass spectrometry. Table 23.4 lists some other diseases of amino acid metabolism.

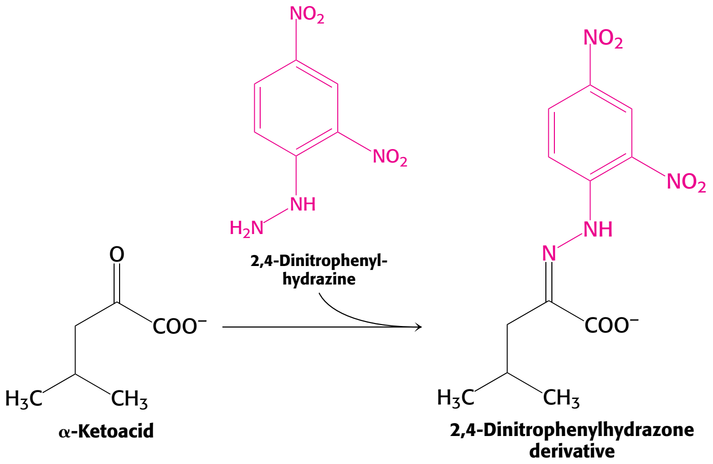
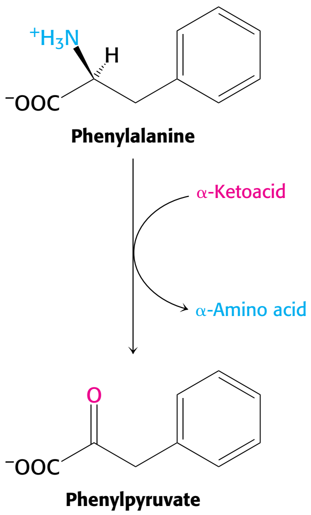

NO₂
NO₂
NH
H₂N
2,4-Dinitrophenyl-hydrazine
O
COO⁻
H₃C
CH₃
α-Ketoacid
NO₂
NO₂
NH
N
COO⁻
H₃C
CH₃
2,4-Dinitrophenylhydrazone derivative

TABLE 23.4 Inborn errors of amino acid metabolism

Disease	Enzyme deficiency	Symptoms
Citrullinema	Arginosuccinate lyase	Lethargy, seizures, reduced muscle tension
Tyrosinemia	Various enzymes of tyrosine degradation	Weakness, liver damage, mental retardation
Albinism	Tyrosinase	Absence of pigmentation
Homocystinuria	Cystathionine β-synthase	Scoliosis, muscle weakness, mental retardation, thin blond hair
Hyperlysinemia	α-Aminoadipic semialdehyde dehydrogenase	Seizures, mental retardation, lack of muscle tone, ataxia

Phenylketonuria is one of the most common metabolic disorders

Phenylketonuria is perhaps the best known of the diseases of amino acid metabolism. Phenylketonuria, which occurs with a prevalence of 1 in 10,000 births, is caused by an *absence or deficiency of phenylalanine hydroxylase* or, more rarely, of its tetrahydrobiopterin cofactor. *Phenylalanine accumulates in all body fluids because it cannot be converted into tyrosine.* Normally, three-quarters of phenylalanine molecules are converted into tyrosine, and the other quarter become incorporated into proteins. Because the major outflow pathway is blocked in phenylketonuria, the blood level of phenylalanine is typically at least 20-fold as high as in normal people. Minor fates of phenylalanine in normal people, such as the formation of phenylpyruvate, become major fates in phenylketonurics. Indeed, the initial description of phenylketonuria in 1934 was made by observing the reaction of phenylpyruvate in the urine of phenylketonurics with $FeCl_3$, which turns the urine olive green.



Almost all untreated phenylketonurics are severely mentally retarded. The brain weight of these people is below normal, myelination of their nerves is defective, and their reflexes are hyperactive. The life expectancy of untreated phenylketonurics is drastically shortened. Half die by age 20 and three-quarters by age 30. Phenylketonurics appear normal at birth but are severely defective by age 1 if untreated. The therapy for phenylketonuria is a *low-phenylalanine diet* supplemented with tyrosine, because tyrosine is normally synthesized from phenylalanine. The aim is to provide just enough phenylalanine to meet the needs for growth and replacement. Proteins that have a low content of phenylalanine, such as casein from milk, are hydrolyzed and phenylalanine is removed by adsorption. A low-phenylalanine diet must be started very soon after birth to prevent irreversible brain damage. In one study, the average IQ of phenylketonurics treated within a few weeks after birth was 93; a control group treated starting at age 1 had an average IQ of 53.

Early diagnosis of phenylketonuria is essential and has been accomplished by mass screening programs of all babies born in the United States and Canada. The phenylalanine level in the blood is the preferred diagnostic criterion because it is more sensitive and reliable than the $FeCl_3$ test. Prenatal diagnosis of phenylketonuria with DNA probes has become feasible because the gene has been cloned and the exact locations of many mutations have been discovered in the protein.

Determining the basis of the neurological symptoms of phenylketonuria is an active area of research

The biochemical basis of retardation is not firmly established, but one hypothesis suggests that the lack of hydroxylase reduces the amount of tyrosine, an important precursor to neurotransmitters such as dopamine. Moreover, high concentrations of phenylalanine prevent amino acid transport of any tyrosine present as well as tryptophan, a precursor to the neurotransmitter serotonin, into the brain. Because all three of the amino acids are transported by the same carrier, phenylalanine will saturate the carrier, preventing access to tyrosine and tryptophan. Finally, high blood levels of phenylalanine result in higher levels of phenylalanine in the brain, and evidence suggests this elevated concentration inhibits glycolysis at pyruvate kinase, disrupts myelination of nerve fibers and reduces the synthesis of several neurotransmitters.

SUMMARY

23.1 Proteins Are Degraded to Amino Acids

Dietary protein is digested in the intestine, producing amino acids that are transported throughout the body. Cellular proteins are degraded at widely variable rates, ranging from minutes to the life of the organism.

23.2 Protein Turnover Is Tightly Regulated

The turnover of cellular proteins is a regulated process requiring complex enzyme systems. Proteins to be degraded are conjugated with ubiquitin, a small conserved protein, in a reaction driven by ATP hydrolysis. The ubiquitin-conjugating system is composed of three distinct enzymes. A large, barrel-shaped complex called the proteasome digests the ubiquitinated proteins. The proteasome also requires ATP hydrolysis to function. The resulting amino acids provide a source of precursors for protein, nucleotide bases, and other nitrogenous compounds.

23.3 The First Step in Amino Acid Degradation Is the Removal of Nitrogen

Surplus amino acids are used as building blocks and as metabolic fuel. The first step in their degradation is the removal of their α-amino groups by transamination to α-ketoacids. Pyridoxal phosphate is the coenzyme in all aminotransferases and in many other enzymes catalyzing amino acid transformations. The α-amino group funnels into α-ketoglutarate to form glutamate, which is then oxidatively deaminated by glutamate dehydrogenase to give NH_4^+ and α-ketoglutarate. NAD^+ or $NADP^+$ is the electron acceptor in this reaction.

23.4 Ammonium Ion Is Converted into Urea in Most Terrestrial Vertebrates

The first step in the synthesis of urea is the formation of carbamoyl phosphate, which is synthesized from HCO_3^-, NH_3, and two molecules of ATP by carbamoyl phosphate synthetase. Ornithine is then carbamoylated to citrulline by ornithine transcarbamoylase. These two reactions take place in mitochondria. Citrulline leaves the mitochondrion and condenses with aspartate to form argininosuccinate, which is cleaved into arginine and fumarate. The other nitrogen atom of urea comes from aspartate. Urea is formed by the hydrolysis of arginine, which also regenerates ornithine.

23.5 Carbon Atoms of Degraded Amino Acids Emerge as Major Metabolic Intermediates

The carbon atoms of degraded amino acids are converted into pyruvate, acetyl CoA, acetoacetate, or an intermediate of the citric acid cycle. Most amino acids are solely glucogenic, two are solely ketogenic, and a few are both ketogenic and glucogenic. Alanine, serine, cysteine, glycine, threonine, and tryptophan are degraded to pyruvate. Asparagine and aspartate are converted into oxaloacetate. α-Ketoglutarate is the point of entry for glutamate and four amino acids (glutamine, histidine, proline, and arginine) that can be converted into glutamate. Succinyl CoA is the point of entry for some of the carbon atoms of three amino acids (methionine, isoleucine, and valine) that are degraded through the intermediate methylmalonyl CoA. Leucine is degraded to acetoacetate and acetyl CoA. The breakdown of valine and isoleucine is like that of leucine. Their α-ketoacid derivatives are oxidatively decarboxylated by the branched-chain α-ketoacid dehydrogenase.

The rings of aromatic amino acids are degraded by oxygenases. Phenylalanine hydroxylase, a monooxygenase, uses tetrahydrobiopterin as the reductant. One of the oxygen atoms of O_2 emerges in tyrosine and the other in water. Subsequent steps in the degradation of these aromatic amino acids are catalyzed by dioxygenases, which catalyze the insertion of both atoms of O_2 into organic products. Four of the carbon atoms of phenylalanine and tyrosine are converted into fumarate, and four emerge in acetoacetate.

23.6 Inborn Errors of Metabolism Can Disrupt Amino Acid Degradation

Errors in amino acid metabolism were sources of some of the first insights into the correlation between pathology and biochemistry. Phenylketonuria is the best known of the many hereditary errors of amino acid metabolism. This condition results from the accumulation of high levels of phenylalanine in the body fluids. This accumulation leads to mental retardation unless the afflicted are placed on low-phenylalanine diets immediately after birth.

KEY TERMS

ubiquitin (p. 683)
degron (p. 684)
proteasome (p. 685)
aminotransferase (transaminase) (p. 688)
glutamate dehydrogenase (p. 688)
pyridoxal phosphate (PLP) (p. 689)
pyridoxamine phosphate (PMP) (p. 690)
glucose–alanine cycle (p. 693)
urea cycle (p. 693)
carbamoyl phosphate synthetase (p. 693)
N-acetylglutamate (p. 694)
ketogenic amino acid (p. 698)
glucogenic amino acid (p. 698)
biopterin (p. 703)
phenylketonuria (p. 706)

PROBLEMS

1. *Getting exposure.* Proteins are denatured by acid in the stomach. This denaturation makes them better substrates for proteolysis. Explain why this is the case.

2. *Targeting for destruction.* What are the steps required to attach ubiquitin to a target protein?

3. *Not the dating service.* Match the description on the right with the term on the left.

(a) Pepsin _____
(b) N-terminal rule _____
(c) Ubiquitin _____
(d) PEST sequence _____
(e) Threonine nucleophiles _____
(f) ATP-dependent protein unfolding _____
(g) Proteasome _____
(h) Ubiquitin-activating enzyme _____
(i) Ubiquitin-conjugating enzyme _____
(j) Ubiquitin-ligase _____

1. Requires an adenylate intermediate
2. Marks a protein for destruction
3. 19S regulatory subunit
4. Determines half-life of a protein
5. 20S core
6. Substrate for ligase
7. Stomach proteolytic enzyme
8. Recognizes protein to be degraded
9. Protein degrading machine
10. Pro-Glu-Ser-Thr

4. *Wasted energy?* Protein hydrolysis is an exergonic process, yet the 26S proteasome is dependent on ATP hydrolysis for activity.

(a) Explain why ATP hydrolysis is required by the 26S proteasome.

(b) Small peptides can be hydrolyzed without the expenditure of ATP. How does this information concur with your answer to part *a*?

5. *Keto counterparts.* Name the α-ketoacid that is formed by the transamination of each of the following amino acids:

(a) Alanine
(b) Aspartate
(c) Glutamate
(d) Leucine
(e) Phenylalanine
(f) Tyrosine

6. *A versatile building block.*

(a) Write a balanced equation for the conversion of aspartate into glucose through the intermediate oxaloacetate. Which coenzymes participate in this transformation? (b) Write a balanced equation for the conversion of aspartate into oxaloacetate through the intermediate fumarate.

7. *The benefits of specialization.* The archaeal proteasome contains 14 identical active β subunits, whereas the eukaryotic proteasome has 7 distinct β subunits. What are the potential benefits of having several distinct active subunits?

8. *Propose a structure.* The 19S subunit of the proteasome contains six subunits that are members of the AAA ATPase family. Other members of this large family are associated into homohexamers with sixfold symmetry. Propose a

structure for the AAA ATPases within the 19S proteasome. How might you test and refine your prediction?

9. *Effective electron sinks.* Pyridoxal phosphate stabilizes carbanionic intermediates by serving as an electron sink. Which other prosthetic group catalyzes reactions in this way?

10. *Cooperation.* How do aminotransferases and glutamate dehydrogenase cooperate in the metabolism of the amino group of amino acids?

11. *Taking away the nitrogen.* What amino acids yield citric acid cycle components and glycolysis intermediates when deaminated?

12. *One reaction only.* What amino acids can be deaminated directly?

13. *Useful products.* What are the common features of the breakdown products of the carbon skeletons of amino acids?

14. *Helping hand.* Propose a role for the positively charged guanidinium nitrogen atom in the cleavage of argininosuccinate into arginine and fumarate.

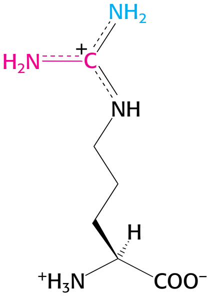
15. *Nitrogen sources.* What are the immediate biochemical sources for the two nitrogen atoms in urea?

16. *Counterparts.* Match the biochemical on the right with the property on the left.

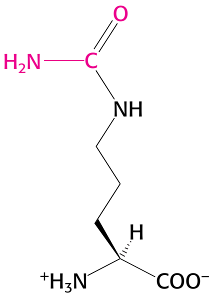
(a) Formed from NH_4^+ _____
(b) Hydrolyzed to yield urea _____
(c) A second source of nitrogen _____
(d) Reacts with aspartate _____
(e) Cleavage yields fumarate _____
(f) Accepts the first nitrogen _____
(g) Final product _____

1. Aspartate
2. Urea
3. Ornithine
4. Carbamoyl phosphate
5. Arginine
6. Citrulline
7. Argininosuccinate

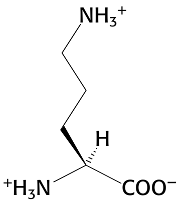
17. *Line up.* Identify structures A–D, and place them in the order that they appear in the urea cycle.



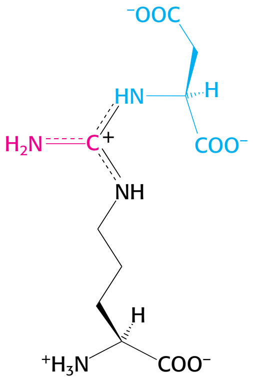
(A)



(B)



(C)

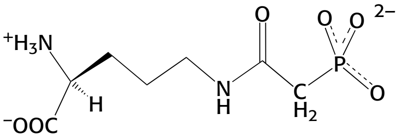


(D)

18. *Completing the cycle.* Four high-transfer-potential phosphoryl groups are consumed in the synthesis of urea according to the stoichiometry given on page 696. In this reaction, aspartate is converted into fumarate. Suppose that fumarate is converted into oxaloacetate. What is the resulting stoichiometry of urea synthesis? How many high-transfer-potential phosphoryl groups are spent?

19. *A good bet.* A friend bets you a bazillion dollars that you can't prove that the urea cycle is linked to the citric acid cycle and other metabolic pathways. Can you collect?

20. *Inhibitor design.* Compound A has been synthesized as a potential inhibitor of a urea-cycle enzyme. Which enzyme do you think compound A might inhibit? Explain your answer.



Compound A

21. *Ammonia toxicity.* Glutamate is an important neurotransmitter whose levels must be carefully regulated in the brain. Explain how a high concentration of ammonia might disrupt this regulation. How might a high concentration of ammonia alter the citric acid cycle?

22. *A precise diagnosis.* The urine of an infant gives a positive reaction with 2,4-dinitrophenylhydrazine. Mass spectrometry shows abnormally high blood levels of pyruvate, α-ketoglutarate, and the α-ketoacids of valine, isoleucine, and leucine. Identify a likely molecular defect and propose a definitive test of your diagnosis.

23. *Therapeutic design.* How would you treat an infant who is deficient in argininosuccinate synthetase? Which molecules would carry nitrogen out of the body?

24. *Damaged liver.* As we will see later (Chapter 27), liver damage (cirrhosis) often results in ammonia poisoning. Explain why this is the case.

25. *Argininosuccinic aciduria*. Argininosuccinic aciduria is a condition that results when the urea-cycle enzyme argininosuccinase is deficient. Argininosuccinate is present in the blood and urine. Suggest how this condition might be treated while still removing nitrogen from the body.

26. *Sweet hazard.* Why should phenylketonurics avoid using aspartame, an artificial sweetener? (Hint: Aspartame is L-aspartyl-L-phenylalanine methyl ester.)

27. *Déjà vu.* *N*-Acetylglutamate is required as a cofactor in the synthesis of carbamoyl phosphate. How is *N*-acetylglutamate synthesized from glutamate?

28. *Negative nitrogen balance.* A deficiency of even one amino acid results in a negative nitrogen balance. In this state, more protein is degraded than is synthesized, and so more nitrogen is excreted than is ingested. Why would protein be degraded if one amino acid were missing?

29. *Precursors.* Differentiate between ketogenic amino acids and glucogenic amino acids.

30. *A slight of hand*. The end products of tryptophan degradation are acetyl CoA and acetoacetyl CoA, yet tryptophan is a gluconeogenic amino acid in animals. Explain.

31. *Closely related*. Pyruvate dehydrogenase complex and α-ketoglutarate dehydrogenase complex are huge enzymes consisting of three discrete enzymatic activities. Which amino acids require a related enzyme complex, and what is the name of the enzyme?

32. *Supply lines*. The carbon skeletons of the 20 common amino acids can be degraded into a limited number of end products. What are the end products and in what metabolic pathway are they commonly found?

Mechanism Problems

33. *Serine dehydratase.* Write out a complete mechanism for the conversion of serine into aminoacrylate catalyzed by serine dehydratase.

34. *Serine racemase.* The nervous system contains a substantial amount of D-serine, which is generated from L-serine by serine racemase, a PLP-dependent enzyme. Propose a mechanism for this reaction. What is the equilibrium constant for the reaction L-serine $\rightleftharpoons$ D-serine?

Chapter Integration Problems

35. *Multiple substrates*. In Chapter 8, we learned that there are two types of bisubstrate reactions, sequential and double-displacement. Which type characterizes the action of aminotransferases? Explain your answer.

36. *Double duty.* Degradation signals are commonly located in protein regions that also facilitate protein–protein interactions. Explain why this coexistence of two functions in the same domain might be useful.

37. *Fuel choice.* Within a few days after a fast begins, nitrogen excretion accelerates to a higher-than-normal level. After a few weeks, the rate of nitrogen excretion falls to a lower level and continues at this low rate. However, after the fat stores have been depleted, nitrogen excretion rises to a high level.

(a) What events trigger the initial surge of nitrogen excretion?

(b) Why does nitrogen excretion fall after several weeks of fasting?

(c) Explain the increase in nitrogen excretion when the lipid stores have been depleted.

38. *A serious situation*. Pyruvate carboxylase deficiency is a fatal disorder. Patients with pyruvate carboxylase deficiency sometimes display some or all of the following symptoms: lactic acidosis, hyperammonemia (excess NH_4^+ in the blood), hypoglycemia, and demyelination of the regions of the brain due to insufficient lipid synthesis. Provide a possible biochemical rationale for each of these observations.

39. *Isoleucine degradation.* Isoleucine is degraded to acetyl CoA and succinyl CoA. Suggest a plausible reaction sequence, based on reactions discussed in the text, for this degradation pathway.

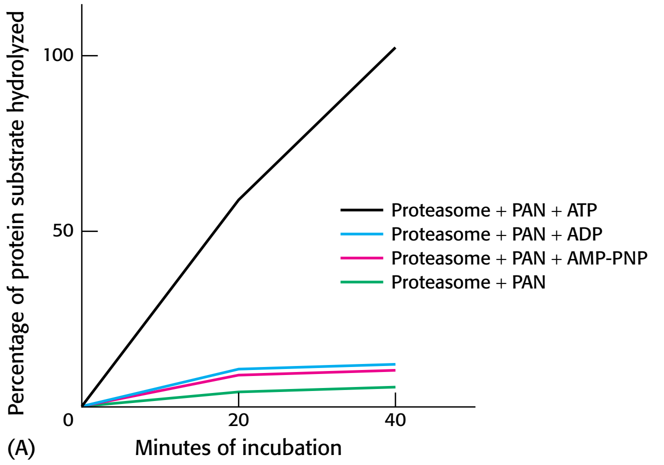
40. *Many roles.* Pyridoxal phosphate is an important coenzyme in transamination reactions. We have seen this coenzyme before, in glycogen metabolism. Which enzyme in glycogen metabolism requires pyridoxal phosphate and what role does the coenzyme play in this enzyme?

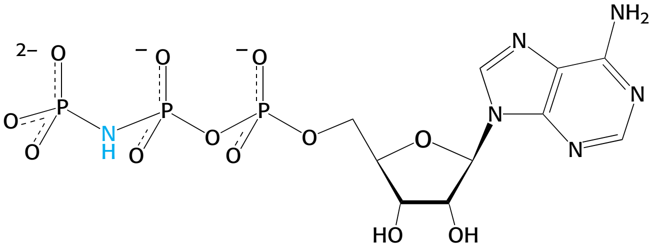
41. *Enough cycles to have a race.* The glucose–alanine cycle is reminiscent of the Cori cycle, but the glucose–alanine cycle can be said to be more energy efficient. Explain why this is so.

Data Interpretation Problem

42. *Another helping hand.* In eukaryotes, the 20S proteasome component in conjunction with the 19S component degrades ubiquitinated proteins with the hydrolysis of a molecule of ATP. Archaea lack ubiquitin and the 26S proteasome but do contain a 20S proteasome. Some archaea also contain an ATPase that is homologous to the ATPases of the eukaryotic 19S component. This archaeal ATPase activity was isolated as a 650-kDa complex (called PAN) from the archaeon *Thermoplasma,* and experiments were performed to determine if PAN could enhance the activity of the 20S proteasome from *Thermoplasma* as well as other 20S proteasomes.

Protein degradation was measured as a function of time and in the presence of various combinations of components. Graph A shows the results.





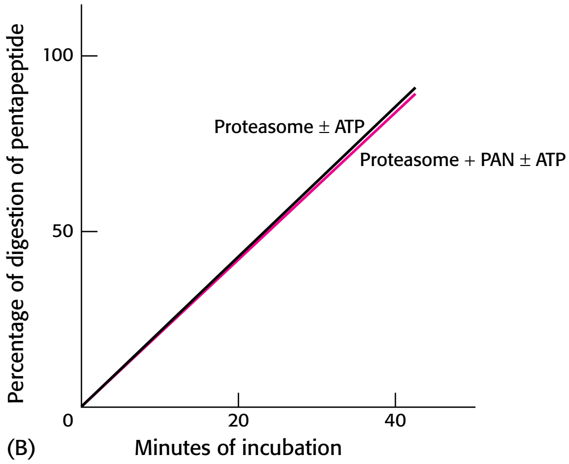
AMP-PNP

(a) What is the effect of PAN on archaeal proteasome activity in the absence of nucleotides?

(b) What is the nucleotide requirement for protein digestion?

(c) What evidence suggests that ATP hydrolysis, and not just the presence of ATP, is required for digestion?

A similar experiment was performed with a small peptide as a substrate for the proteasome instead of a protein. The results obtained are shown in graph B.



(d) How do the requirements for peptide digestion differ from those of protein digestion?

(e) Suggest some reasons for the difference.

The ability of PAN from the archaeon *Thermoplasma* to support protein degradation by the 20S proteasomes from the archaeon *Methanosarcina* and rabbit muscle was then examined.

Percentage of digestion of protein substrate
(20S proteasome source)

Additions	Thermoplasma	Methanosarcina	Rabbit muscle
None	11	10	10
PAN	8	8	8
PAN + ATP	100	40	30
PAN + ADP	12	9	10

[Data from P. Zwickl, D. Ng, K. M. Woo, H.-P. Klenk, and A. L. Goldberg. An archaebacterial ATPase, homologous to ATPase in the eukaryotic 26S proteasome, activates protein breakdown by 20S proteasomes. *J. Biol. Chem.* 274(1999): 26008–26014.]

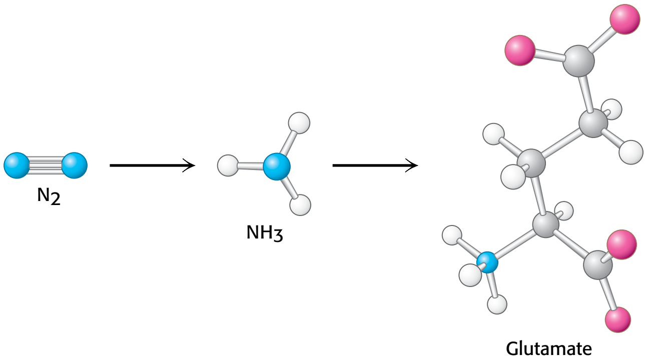
(f) Can the *Thermoplasma* PAN augment protein digestion by the proteasomes from other organisms?

(g) What is the significance of the stimulation of rabbit muscle proteasome by *Thermoplasma* PAN?

CHAPTER

24

The Biosynthesis of Amino Acids



Nitrogen is a key component of amino acids. The atmosphere is rich in nitrogen gas (N_2), a very unreactive molecule. Certain organisms, such as bacteria that live in the root nodules of yellow clover, can convert nitrogen gas into ammonia (NH_3), which can then be used to synthesize, first, glutamate and then other amino acids. [(Left) Hugh Spencer/Science Source.]

The assembly of biological molecules, including proteins and nucleic acids, requires the generation of appropriate starting materials. We have already considered the assembly of carbohydrates in discussions of the Calvin cycle and the pentose phosphate pathway (Chapter 20). The present chapter and the next two examine the assembly of the other important building blocks—namely, amino acids, nucleotides, and lipids.

The pathways for the biosynthesis of these molecules are extremely ancient, going back to the last common ancestor of all living things. Indeed, these pathways probably predate many of the pathways of energy transduction discussed in Part II and may have provided key selective advantages in early evolution. Many of the intermediates in energy-yielding pathways play a role in biosynthesis as well. These common intermediates allow efficient interplay between energy-yielding (catabolic) and energy-requiring biosynthetic (anabolic) pathways. Thus, cells are able to balance the degradation of compounds for energy mobilization and the synthesis of starting materials for macromolecular construction.

We begin our consideration of biosynthesis with amino acids—the building blocks of proteins and the nitrogen source for many other important molecules, including nucleotides, neurotransmitters, and prosthetic groups such as porphyrins. Amino acid biosynthesis is intimately connected with nutrition because many higher organisms, including human beings,

OUTLINE

24.1 Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia

24.2 Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways

24.3 Feedback Inhibition Regulates Amino Acid Biosynthesis

24.4 Amino Acids Are Precursors of Many Biomolecules

Anabolism

Biosynthetic processes.

Catabolism

Degradative processes.

Derived from the Greek *ana,* "up"; *kata,* "down"; *ballein,* "to throw."

have lost the ability to synthesize some amino acids and must therefore obtain adequate quantities of these essential amino acids in their diets. Furthermore, because some amino acid biosynthetic enzymes are absent in mammals but present in plants and microorganisms, they are useful targets for herbicides and antibiotics.

Amino acid synthesis requires solutions to three key biochemical problems

Nitrogen is an essential component of amino acids. Earth has an abundant supply of nitrogen, but it is primarily in the form of atmospheric nitrogen gas (N_2), a remarkably inert molecule. Thus, a fundamental problem for biological systems is to obtain nitrogen in a more usable form. This problem is solved by certain microorganisms capable of reducing the inert $N \equiv N$ molecule of nitrogen gas to two molecules of ammonia in one of the most amazing reactions in biochemistry. Nitrogen in the form of ammonia is the source of nitrogen for all the amino acids. The carbon backbones come from the glycolytic pathway, the pentose phosphate pathway, or the citric acid cycle.

In amino acid production, we encounter an important problem in biosynthesis—namely, stereochemical control. Because all amino acids except glycine are chiral, biosynthetic pathways must generate the correct isomer with high fidelity. In each of the 19 pathways for the generation of chiral amino acids, the stereochemistry at the α-carbon atom is established by a transamination reaction that includes pyridoxal phosphate (PLP). Almost all the aminotransferases that catalyze these reactions descend from a common ancestor, illustrating once again that effective solutions to biochemical problems are retained throughout evolution.

Biosynthetic pathways are often highly regulated such that building blocks are synthesized only when supplies are low. Very often, a high concentration of the final product of a pathway inhibits the activity of allosteric enzymes that function early in the pathway to control the committed step. These enzymes are similar in functional properties to aspartate transcarbamoylase and its regulators (Section 10.1). Feedback and allosteric mechanisms ensure that all 20 amino acids are maintained in sufficient amounts for protein synthesis and other processes.

24.1 Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia

The nitrogen in amino acids, purines, pyrimidines, and other biomolecules ultimately comes from atmospheric nitrogen, N_2. The biosynthetic process starts with the reduction of N_2 to NH_3 (ammonia), a process called *nitrogen fixation.* The extremely strong $N \equiv N$ bond, which has a bond energy of 940 kJ mol^{-1} (225 kcal mol^{-1}), is highly resistant to chemical attack. Indeed, Antoine Lavoisier named nitrogen gas "azote," from Greek words meaning "without life," because it is so unreactive. Nevertheless, the conversion of nitrogen and hydrogen to form ammonia is thermodynamically favorable; the reaction is difficult kinetically because of the activation energy required to form intermediates along the reaction pathway.

Although higher organisms are unable to fix nitrogen, this conversion is carried out by some bacteria and archaea. Symbiotic *Rhizobium* bacteria invade the roots of leguminous plants and form root nodules in which they fix nitrogen, supplying both the bacteria and the plants. The importance of nitrogen fixation by *diazotrophic (nitrogen-fixing) microorganisms* to the metabolism of all higher eukaryotes cannot be overstated: the amount of N_2 fixed by these species has been estimated to be 10^{11} kilograms per year, about 60% of Earth's newly fixed nitrogen. Lightning and ultraviolet radiation fix another 15%; the other 25% is fixed by industrial processes. The industrial process for nitrogen fixation devised by Fritz Haber in 1910 is still being used in fertilizer factories.

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$

The fixation of N_2 is typically carried out by mixing with H_2 gas over an iron catalyst at about 500°C and a pressure of 300 atmospheres.

To meet the kinetic challenge, the biological process of nitrogen fixation requires a complex enzyme with multiple redox centers. The *nitrogenase complex,* which carries out this fundamental transformation, consists of two proteins: a *reductase* (also called the iron protein or Fe protein), which provides electrons with high reducing power, and *nitrogenase* (also called the molybdenum–iron protein or MoFe protein), which uses these electrons to reduce N_2 to NH_3. The transfer of electrons from the reductase to the nitrogenase component is coupled to the hydrolysis of ATP by the reductase (Figure 24.1).

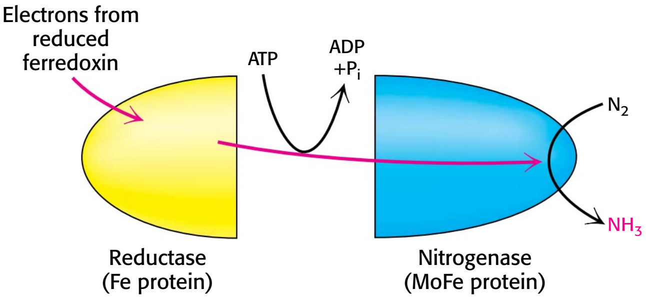


FIGURE 24.1 Nitrogen fixation. Electrons flow from ferredoxin to the reductase (iron protein, or Fe protein) to nitrogenase (molybdenum–iron protein, or MoFe protein) to reduce nitrogen to ammonia. ATP hydrolysis within the reductase drives conformational changes necessary for the efficient transfer of electrons.

In principle, the reduction of N_2 to NH_3 is a six-electron process.

$$N_2 + 6\,e^- + 6\,H^+ \longrightarrow 2\,NH_3$$

However, the biological reaction always generates at least 1 mol of H_2 in addition to 2 mol of NH_3 for each mol of $N \equiv N$. Hence, an input of two additional electrons is required.

$$N_2 + 8\,e^- + 8\,H^+ \longrightarrow 2\,NH_3 + H_2$$

In most nitrogen-fixing microorganisms, *the eight high-potential electrons come from reduced ferredoxin,* generated by oxidative processes. Two molecules of ATP are hydrolyzed for each electron transferred. Thus, *at least 16 molecules of ATP are hydrolyzed for each molecule of N_2 reduced.*

$$N_2 + 8\,e^- + 8\,H^+ + 16\,ATP + 16\,H_2O \longrightarrow$$
$$2\,NH_3 + H_2 + 16\,ADP + 16\,P_i$$

Note that O_2 is required for oxidative phosphorylation to generate the ATP necessary for nitrogen fixation. However, the nitrogenase complex is exquisitely sensitive to inactivation by O_2. To allow ATP synthesis and nitrogenase to function simultaneously, leguminous plants maintain a very low concentration of free O_2 in their root nodules, the location of the nitrogenase. This is accomplished by binding O_2 to *leghemoglobin,* a homolog of hemoglobin (Section 6.3).

The iron–molybdenum cofactor of nitrogenase binds and reduces atmospheric nitrogen

Both the reductase and the nitrogenase components of the complex are *iron–sulfur proteins,* in which iron is bonded to the sulfur atom of a cysteine residue and to inorganic sulfide. Recall that iron–sulfur clusters act as electron carriers (Section 18.3). The *reductase* is a dimer of identical 30-kDa subunits bridged by a 4Fe–4S cluster (Figure 24.2).

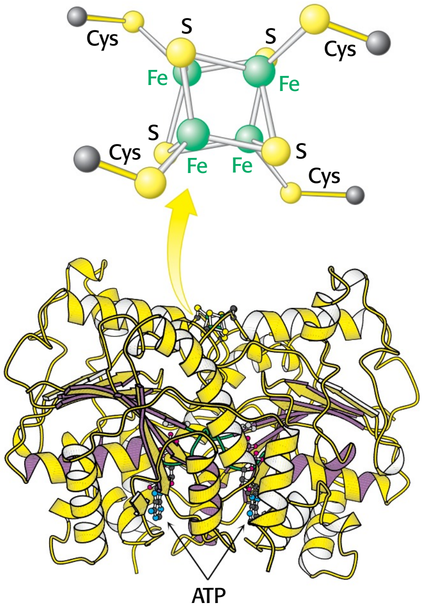


FIGURE 24.2 Fe Protein. This protein is a dimer composed of two polypeptide chains linked by a 4Fe–4S cluster. *Notice* that each monomer is a member of the P-loop NTPase family and contains an ATP-binding site. [Drawn from 1N2C.pdb.]

The role of the reductase is to transfer electrons from a suitable donor, such as reduced ferredoxin, to the nitrogenase component. The 4Fe–4S cluster carries the electrons, one at a time, to nitrogenase. The binding and hydrolysis of ATP triggers a conformational change that moves the reductase closer to the nitrogenase component, whence it is able to transfer its electron to the center of nitrogen reduction. The structure of the ATP-binding region reveals it to be a member of the P-loop NTPase family (Section 9.4) that is clearly related to the nucleotide-binding regions found in G proteins and related proteins. Thus, we see another example of how this domain has been recruited in evolution because of its ability to couple nucleoside triphosphate hydrolysis to conformational changes.

The nitrogenase component is an $\alpha_2\beta_2$ tetramer (240 kDa), in which the α and β subunits are homologous to each other and structurally quite similar (Figure 24.3). The nitrogenase requires the FeMo cofactor, which consists of [Fe_4–S_3] and [Mo–Fe_3–S_3] subclusters joined by three disulfide bonds. A carbon atom (the interstitial carbon), donated by S-adenosylmethionine (p. 724), sits at the interstices of the iron atoms of the FeMo cofactor. The FeMo cofactor is also coordinated to a homocitrate moiety and to the α subunit through one histidine residue and one cysteinate residue.

Electrons from the reductase enter at the *P clusters,* which are located at the α–β interface. The role of the P clusters is to store electrons until they can be used productively to reduce nitrogen at the FeMo cofactor. *The FeMo cofactor is the site of nitrogen fixation.* One face of the FeMo cofactor is likely to be the site of nitrogen reduction. The electron-transfer reactions from the P cluster take place in concert with the binding of hydrogen ions to nitrogen as it is reduced. Further studies are under way to elucidate the mechanism of this remarkable reaction.

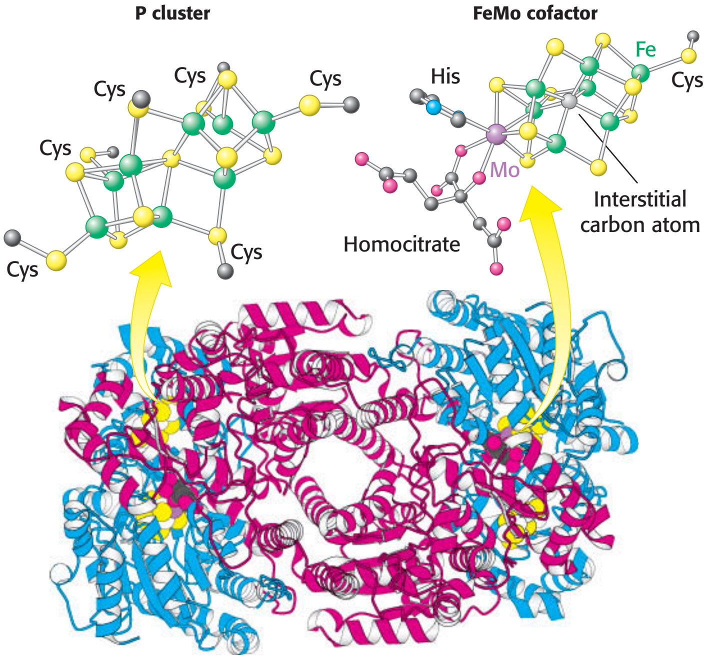
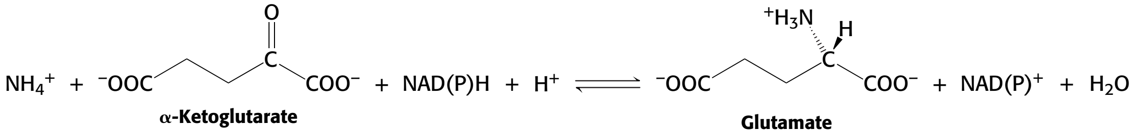


FIGURE 24.3 MoFe protein. This protein is a heterotetramer composed of two α subunits (red) and two β subunits (blue). *Notice* that the protein contains two copies each of two types of clusters: P clusters and FeMo cofactors. Each P cluster contains eight iron atoms (green) and seven sulfides linked to the protein by six cysteinate residues. Each FeMo cofactor contains one molybdenum atom, seven iron atoms, nine sulfides, the interstitial carbon atom, and a homocitrate, and is linked to the protein by one cysteinate residue and one histidine residue. [Drawn from 1M1N.pdb.]

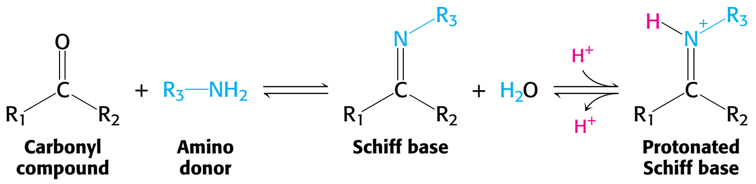
Ammonium ion is assimilated into an amino acid through glutamate and glutamine

The next step in the assimilation of nitrogen into biomolecules is the entry of NH_4^+ into amino acids. The amino acids *glutamate* and *glutamine* play pivotal roles in this regard, acting as nitrogen donors for most amino acids. The α-amino group of most amino acids comes from the α-amino group of glutamate by transamination (Section 23.3). Glutamine, the other major nitrogen donor, contributes its side-chain nitrogen atom in the biosynthesis of a wide range of important compounds, including the amino acids tryptophan and histidine.

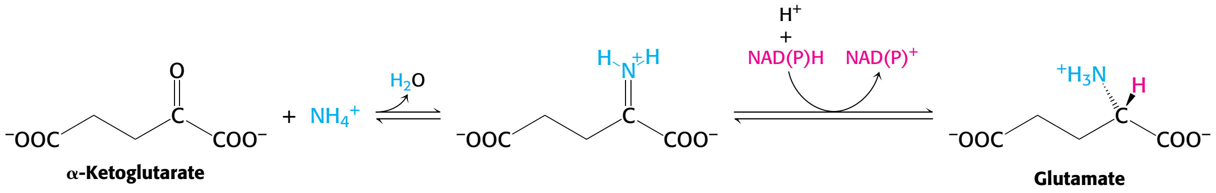
Glutamate is synthesized from NH_4^+ and α-ketoglutarate, a citric acid cycle intermediate, by the action of *glutamate dehydrogenase*. We have already encountered this enzyme in the degradation of amino acids (Section 23.3). Recall that NAD^+ is the oxidant in catabolism, whereas NADPH is the reductant in biosyntheses. Glutamate dehydrogenase is unusual in that it does not discriminate between NADH and NADPH, at least in some species.



The reaction proceeds in two steps. First, a Schiff base forms between ammonia and α-ketoglutarate. The formation of a Schiff base between an amine and a carbonyl compound is a key reaction that takes place at many stages of amino acid biosynthesis and degradation.



Schiff bases are easily protonated. In the second step, the protonated Schiff base is reduced by the transfer of a hydride ion from NAD(P)H to form glutamate.



This reaction is crucial because it establishes the stereochemistry of the α-carbon atom (*S* absolute configuration) in glutamate. The enzyme binds the α-ketoglutarate substrate in such a way that hydride transferred from NAD(P)H is added to form the L isomer of glutamate (Figure 24.4). As we shall see, this stereochemistry is established for other amino acids by transamination reactions that rely on pyridoxal phosphate.

A second ammonium ion is incorporated into glutamate to form glutamine by the action of *glutamine synthetase*. This amidation is driven by the hydrolysis of ATP. ATP participates directly in the reaction by phosphorylating the

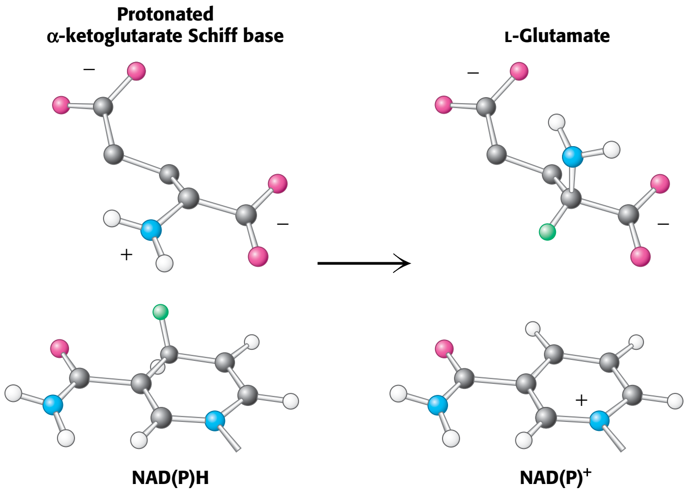
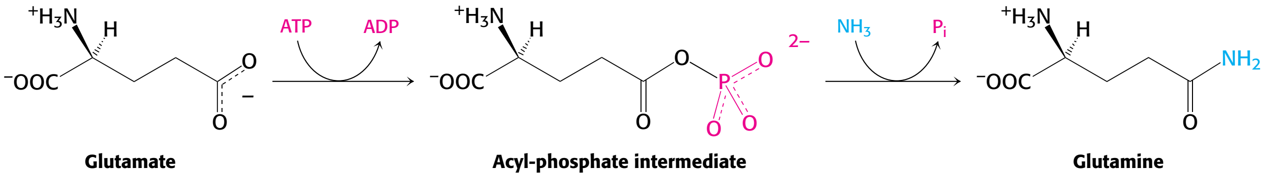


FIGURE 24.4 Establishment of chirality. In the active site of glutamate dehydrogenase, hydride transfer (green) from NAD(P)H to a specific face of the achiral protonated Schiff base of α-ketoglutarate establishes the L configuration of glutamate.

side chain of glutamate to form an acylphosphate intermediate, which then reacts with ammonia to form glutamine.



A high-affinity ammonia-binding site is formed in the enzyme only after the formation of the acylphosphate intermediate. A specific site for ammonia binding is required to prevent attack by water from hydrolyzing the intermediate and wasting a molecule of ATP. The regulation of glutamine synthetase plays a critical role in controlling nitrogen metabolism (Section 24.3).

Glutamate dehydrogenase and glutamine synthetase are present in all organisms. Most prokaryotes also contain an evolutionarily unrelated enzyme, *glutamate synthase,* which catalyzes the reductive amination of α-ketoglutarate to glutamate. Glutamine is the nitrogen donor.

$$\alpha\text{-Ketoglutarate} + \text{glutamine} + \text{NADPH} + \text{H}^+ \rightleftharpoons 2\ \text{glutamate} + \text{NADP}^+$$

The side-chain amide of glutamine is hydrolyzed to generate ammonia within the enzyme, a recurring theme throughout nitrogen metabolism. *When NH_4^+ is limiting, most of the glutamate is made by the sequential action of glutamine synthetase and glutamate synthase.* The sum of these reactions is

$$\text{NH}_4^+ + \alpha\text{-ketoglutarate} + \text{NADPH} + \text{ATP} \longrightarrow \text{glutamate} + \text{NADP}^+ + \text{ADP} + \text{P}_i$$

Note that this stoichiometry differs from that of the glutamate dehydrogenase reaction in that ATP is hydrolyzed. Why do prokaryotes sometimes use this more expensive pathway? The answer is that the value of K_M of glutamate dehydrogenase for NH_4^+ is high (~1 mM), and so this enzyme is not saturated when NH_4^+ is limiting. In contrast, glutamine synthetase has very low K_M for NH_4^+. Thus, ATP hydrolysis is required to capture ammonia when it is scarce.

24.2 Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways

Thus far, we have considered the conversion of N_2 into NH_4^+ and the assimilation of NH_4^+ into glutamate and glutamine. We turn now to the biosynthesis of the other amino acids, the majority of which obtain their nitrogen from glutamate or glutamine. The pathways for the biosynthesis of amino acids are diverse. However, they have an important common feature: *their carbon skeletons come from intermediates of glycolysis, the pentose phosphate pathway, or the citric acid cycle.* On the basis of these starting materials, amino acids can be grouped into six biosynthetic families (Figure 24.5).

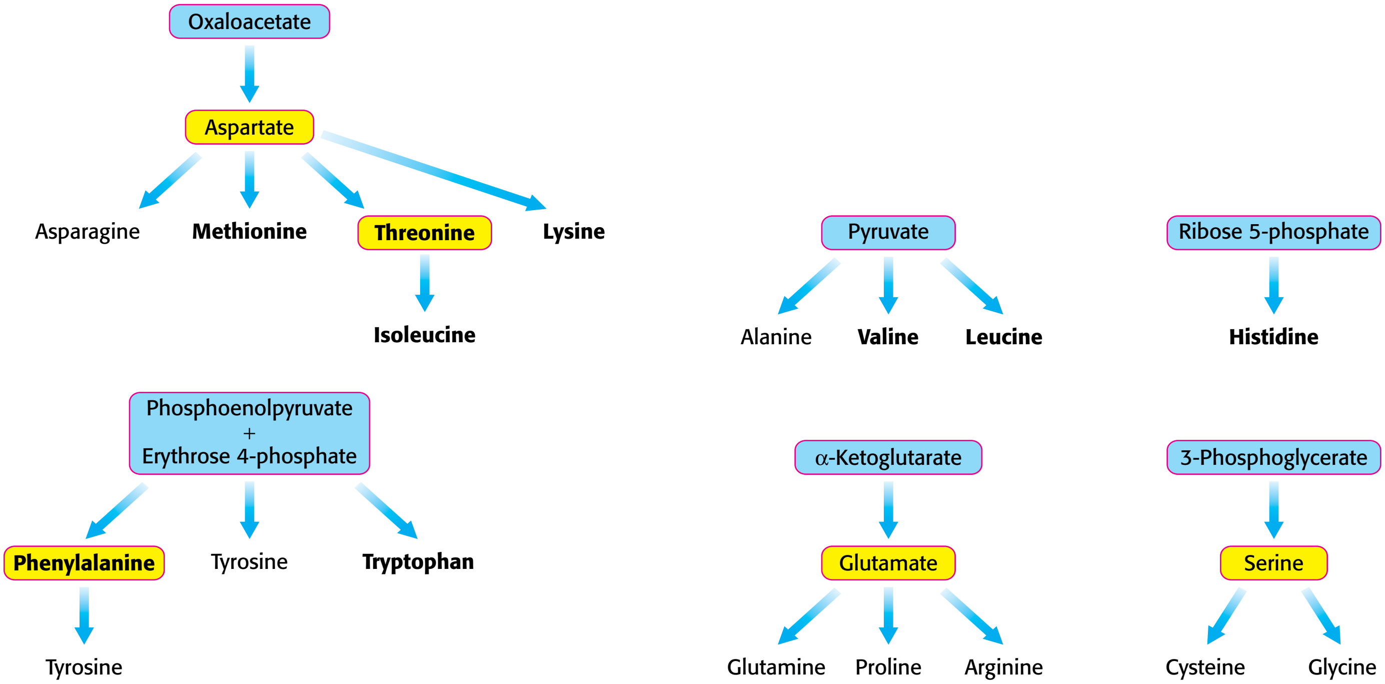

Oxaloacetate
Aspartate
Asparagine
Methionine
Threonine
Lysine
Isoleucine
Pyruvate
Alanine
Valine
Leucine
Ribose 5-phosphate
Histidine
Phosphoenolpyruvate
+
Erythrose 4-phosphate
Phenylalanine
Tyrosine
Tryptophan
Tyrosine
α-Ketoglutarate
Glutamate
Glutamine
Proline
Arginine
3-Phosphoglycerate
Serine
Cysteine
Glycine

FIGURE 24.5 **Biosynthetic families of amino acids in bacteria and plants.** Major metabolic precursors are shaded blue. Amino acids that give rise to other amino acids are shaded yellow. Essential amino acids are in boldface type.

Human beings can synthesize some amino acids but must obtain others from their diet

Most microorganisms, such as *E. coli,* can synthesize the entire basic set of 20 amino acids, whereas human beings cannot make 9 of them. The amino acids that *must* be supplied in the diet are called *essential amino acids,* whereas the others, which can be synthesized if dietary content is insufficient, are termed *nonessential amino acids* (Table 24.1). These designations refer to the needs of an organism under a particular set of conditions. For example, enough arginine is synthesized by the urea cycle to meet the needs of an adult, but perhaps not those of a growing child. A deficiency of even one amino acid results in a *negative nitrogen balance.* In this state, more protein is degraded than is synthesized, and so more nitrogen is excreted than is ingested.

The nonessential amino acids are synthesized by quite simple reactions, whereas the pathways for the formation of the essential amino acids are quite complex. For example, the nonessential amino acids *alanine* and *aspartate* are synthesized in a single step from pyruvate and oxaloacetate, respectively. In contrast, the pathways for the essential amino acids require

TABLE 24.1 Basic set of 20 amino acids

Nonessential	Essential
Alanine	Histidine
Arginine	Isoleucine
Asparagine	Leucine
Aspartate	Lysine
Cysteine	Methionine
Glutamate	Phenylalanine
Glutamine	Threonine
Glycine	Tryptophan
Proline	Valine
Serine	
Tyrosine	

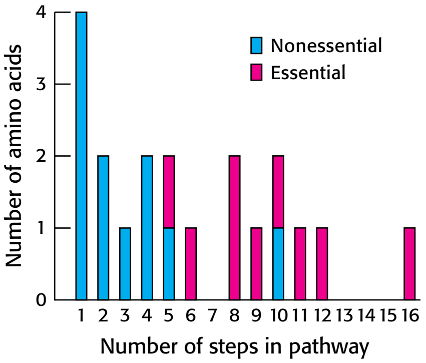


FIGURE 24.6 Essential and nonessential amino acids. Some amino acids are nonessential to human beings because they can be biosynthesized in a small number of steps. Those amino acids requiring a large number of steps for their synthesis are essential in the diet because some of the enzymes for these steps have been lost in the course of evolution.

from 5 to 16 steps (Figure 24.6). The sole exception to this pattern is arginine, inasmuch as the synthesis of this nonessential amino acid de novo requires 10 steps. Typically, however, it is made in only 3 steps from ornithine as part of the urea cycle. Tyrosine, classified as a nonessential amino acid because it can be synthesized in 1 step from phenylalanine, requires 10 steps to be synthesized from scratch and is essential if phenylalanine is not abundant. We begin with the biosynthesis of nonessential amino acids.

Aspartate, alanine, and glutamate are formed by the addition of an amino group to an alpha-ketoacid

Three α-ketoacids—α-ketoglutarate, oxaloacetate, and pyruvate—can be converted into amino acids in one step through the addition of an amino group. We have seen that α-ketoglutarate can be converted into glutamate by reductive amination (Section 23.3). The amino group from glutamate can be transferred to other α-ketoacids by transamination reactions. Thus, aspartate and alanine can be made from the addition of an amino group to oxaloacetate and pyruvate, respectively.

$$\text{Oxaloacetate} + \text{glutamate} \rightleftharpoons \text{aspartate} + \alpha\text{-ketoglutarate}$$
$$\text{Pyruvate} + \text{glutamate} \rightleftharpoons \text{alanine} + \alpha\text{-ketoglutarate}$$

These reactions are carried out by *pyridoxal phosphate-dependent aminotransferases.* Transamination reactions are required for the synthesis of most amino acids.

In Section 23.3, we considered the mechanism of aminotransferases as applied to the metabolism of amino acids. Let us review the aminotransferase mechanism as it operates in the *biosynthesis* of amino acids (Figure 23.11). The reaction pathway begins with *pyridoxal phosphate* in a Schiff-base linkage with lysine at the aminotransferase active site, forming an internal aldimine (Figure 24.7). An amino group is transferred from glutamate to form pyridoxamine phosphate (PMP), the actual amino donor, in a multistep process. PMP then reacts with an incoming α-ketoacid to form a ketimine. Proton loss

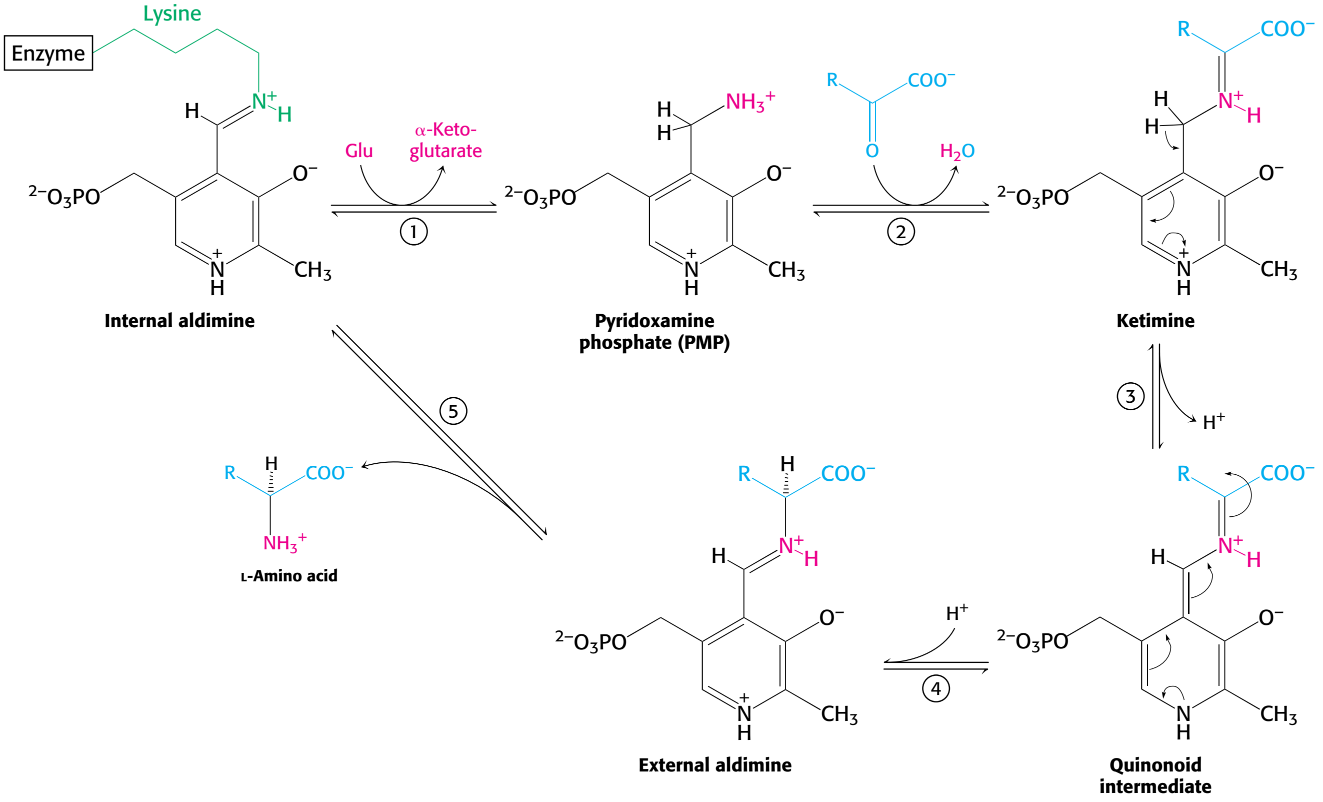


FIGURE 24.7 Amino acid biosynthesis by transamination. (1) Within aminotransferase, the internal aldimine is converted into pyridoxamine phosphate (PMP) by reaction with glutamate in a multistep process not shown. (2) PMP then reacts with an α-ketoacid to generate a ketimine. (3) This intermediate is converted into a quinonoid intermediate (4), which in turn yields an external aldimine. (5) The aldimine is cleaved to release the newly formed amino acid to complete the cycle.

forms a quinonoid intermediate that then accepts a proton at a different site to form an external aldimine. The newly formed amino acid is released with the concomitant formation of the internal aldimine.

A common step determines the chirality of all amino acids

Aspartate aminotransferase is the prototype of a large family of PLP-dependent enzymes. Comparisons of amino acid sequences as well as several three-dimensional structures reveal that almost all aminotransferases having roles in amino acid biosynthesis are related to aspartate aminotransferase by divergent evolution. An examination of the aligned amino acid sequences reveals that two residues are completely conserved. These residues are the lysine residue that forms the Schiff base with the PLP cofactor (lysine 258 in aspartate aminotransferase) and an arginine residue that interacts with the α-carboxylate group of the ketoacid (Figure 23.12).

An essential step in the transamination reaction is the protonation of the quinonoid intermediate to form the external aldimine. *The chirality of the amino acid formed is determined by the direction from which this proton is added to the quinonoid form* (Figure 24.8). The interaction between the conserved arginine residue and the α-carboxylate group helps orient the substrate so that the lysine residue transfers a proton to the bottom face of the quinonoid intermediate, generating an aldimine with an L configuration at the C_α center.

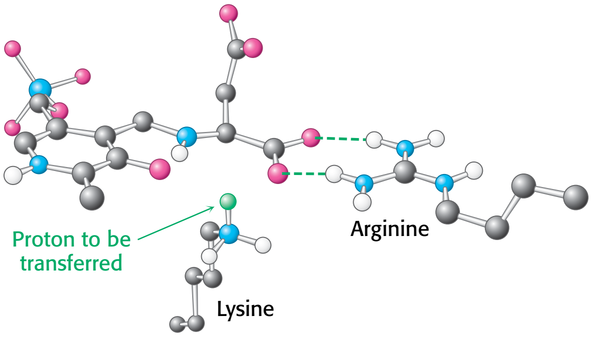


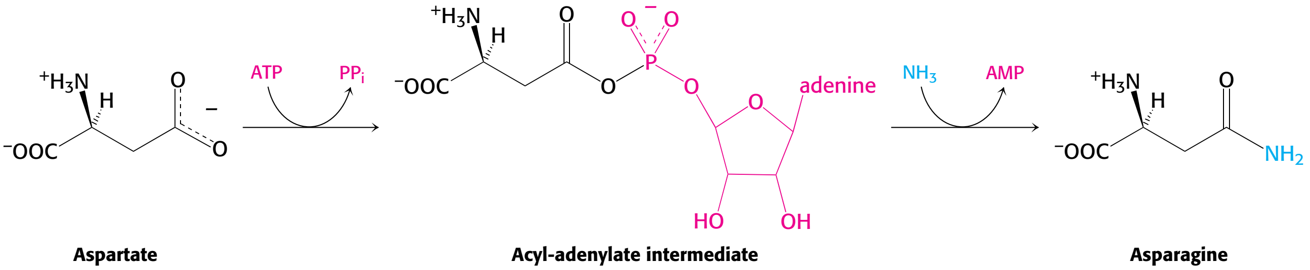
FIGURE 24.8 Stereochemistry of proton addition. In an aminotransferase active site, the addition of a proton from the lysine residue to the bottom face of the quinonoid intermediate determines the L configuration of the amino acid product. The conserved arginine residue interacts with the α-carboxylate group and helps establish the appropriate geometry of the quinonoid intermediate.

The formation of asparagine from aspartate requires an adenylated intermediate

The formation of asparagine from aspartate is chemically analogous to the formation of glutamine from glutamate. Both transformations are amidation reactions and both are driven by the hydrolysis of ATP. The actual reactions are different, however. In bacteria, the reaction for the asparagine synthesis is

$$\text{Aspartate} + NH_4^+ + ATP \longrightarrow \text{asparagine} + AMP + PP_i + H^+$$

Thus, the products of ATP hydrolysis are AMP and PP_i rather than ADP and P_i. Aspartate is activated by adenylation rather than by phosphorylation.

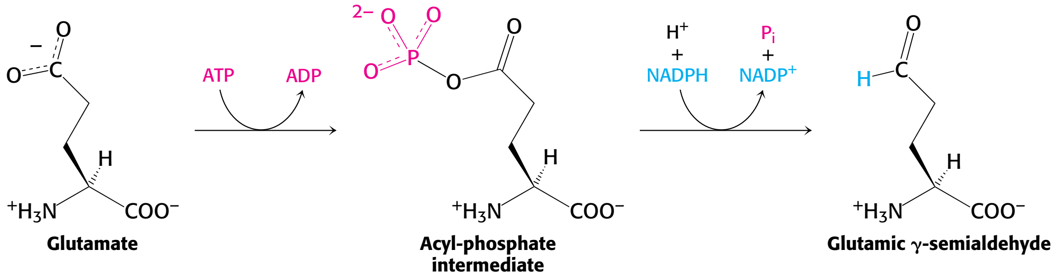


We have encountered this mode of activation in fatty acid degradation and will see it again in lipid and protein synthesis.

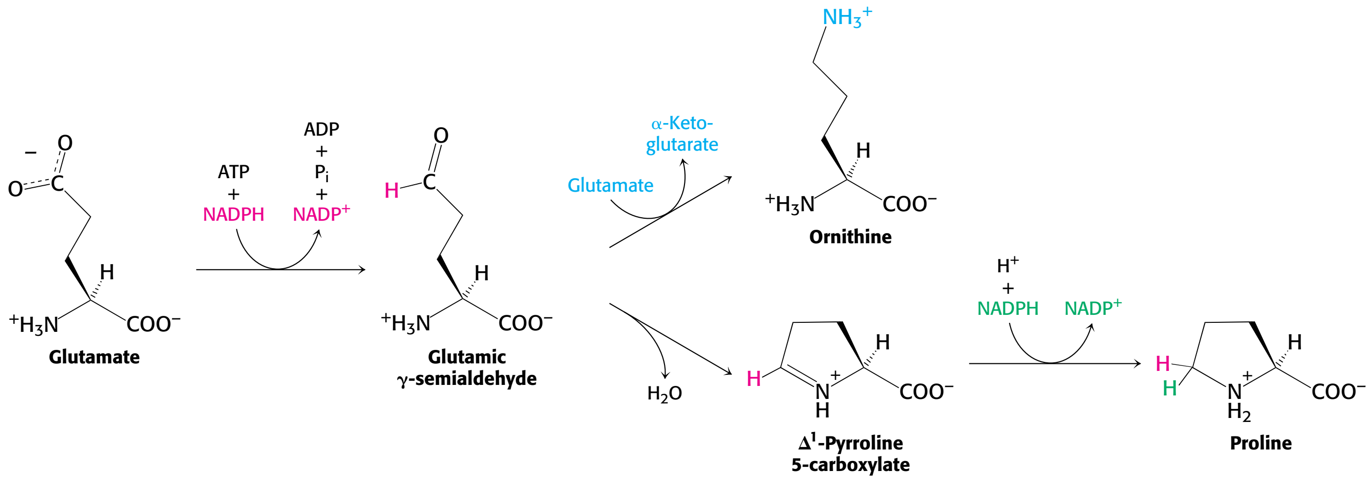
In mammals, the nitrogen donor for asparagine is glutamine rather than ammonia as in bacteria. Ammonia is generated by hydrolysis of the side chain of glutamine and directly transferred to activated aspartate, bound in the active site. An advantage is that the cell is not directly exposed to NH_4^+, which is toxic at high levels to human beings and other mammals. *The use of glutamine hydrolysis as a mechanism for generating ammonia for use within the same enzyme is a motif common throughout biosynthetic pathways.*

Glutamate is the precursor of glutamine, proline, and arginine

The synthesis of glutamate by the reductive amination of α-ketoglutarate has already been discussed, as has the conversion of glutamate into glutamine (p. 717). Glutamate is the precursor of two other nonessential amino acids: *proline* and *arginine*. First, the γ-carboxyl group of glutamate reacts with ATP to form an acyl phosphate. This mixed anhydride is then reduced by NADPH to an aldehyde.

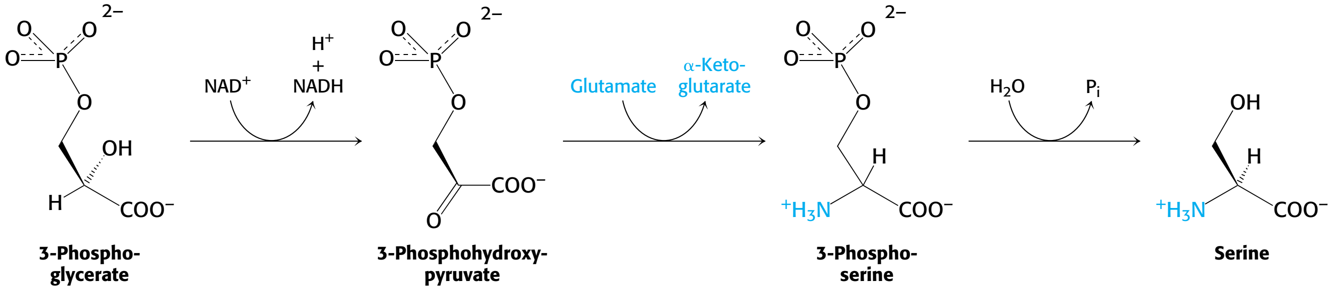

Glutamate
ATP
ADP
Acyl-phosphate intermediate
H⁺ + NADPH
Pᵢ + NADP⁺
Glutamic γ-semialdehyde

Glutamic γ-semialdehyde cyclizes with a loss of H_2O in a nonenzymatic process to give Δ^1-pyrroline 5-carboxylate, which is reduced by NADPH to proline. Alternatively, the semialdehyde can be transaminated to ornithine, which is converted in several steps into arginine in the urea cycle (Figure 23.17).


Glutamate
ATP + NADPH
ADP + Pᵢ + NADP⁺
Glutamic γ-semialdehyde
Glutamate
α-Keto-glutarate
Ornithine
H₂O
Δ¹-Pyrroline 5-carboxylate
H⁺ + NADPH
NADP⁺
Proline

3-Phosphoglycerate is the precursor of serine, cysteine, and glycine

Serine is synthesized from 3-phosphoglycerate, an intermediate in glycolysis. The first step is an oxidation to 3-phosphohydroxypyruvate. This α-ketoacid is transaminated to 3-phosphoserine, which is then hydrolyzed to serine.


3-Phospho-glycerate
NAD⁺
H⁺ + NADH
3-Phosphohydroxy-pyruvate
Glutamate
α-Keto-glutarate
3-Phospho-serine
H₂O
Pᵢ
Serine

Serine is the precursor of *cysteine* and *glycine*. As we shall see, the conversion of serine into cysteine requires the substitution of a sulfur atom derived from methionine for the side-chain oxygen atom. In the formation of glycine,

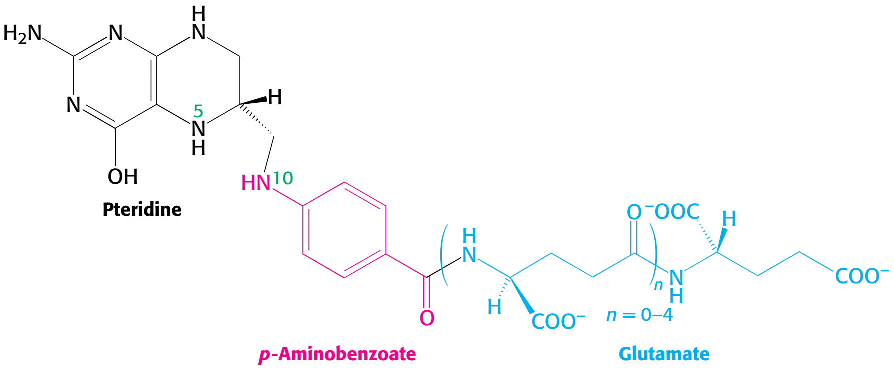
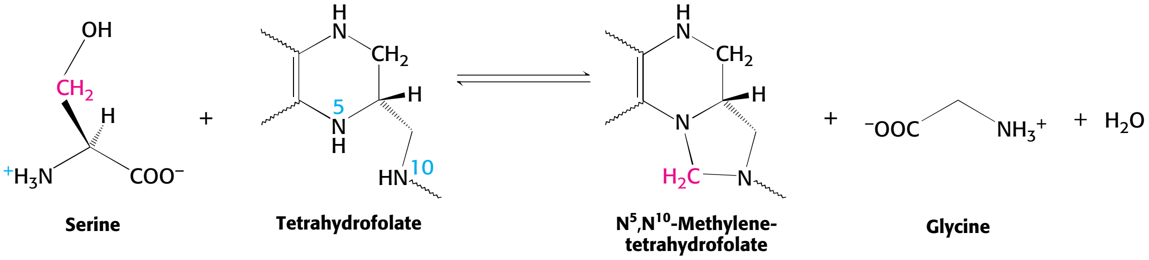


FIGURE 24.9 Tetrahydrofolate. This cofactor includes three components: a pteridine ring, *p*-aminobenzoate, and one or more glutamate residues.

the side-chain methylene group of serine is transferred to *tetrahydrofolate,* a carrier of one-carbon units that will be discussed shortly (Figure 24.9).



This interconversion is catalyzed by *serine hydroxymethyltransferase,* a PLP enzyme that is homologous to aspartate aminotransferase. The formation of the Schiff base of serine renders the bond between its α- and β-carbon atoms susceptible to cleavage, enabling the transfer of the β-carbon to tetrahydrofolate and producing the Schiff base of glycine.

Tetrahydrofolate carries activated one-carbon units at several oxidation levels

Tetrahydrofolate (also called *tetrahydropteroylglutamate*) is a highly versatile carrier of activated one-carbon units. This cofactor consists of three groups: a substituted pteridine, *p*-aminobenzoate, and a chain of one or more glutamate residues (Figure 24.9). Mammals can synthesize the pteridine ring, but they are unable to conjugate it to the other two units. They obtain tetrahydrofolate from their diets or from microorganisms in their intestinal tracts.

The one-carbon group carried by tetrahydrofolate is bonded to its N-5 or N-10 nitrogen atom (denoted as N^5 and N^{10}) or to both. This unit can exist in three oxidation states (Table 24.2). The most-reduced form carries a *methyl* group, whereas the intermediate form carries a *methylene* group. More-oxidized forms carry a *formyl, formimino,* or *methenyl* group. The fully oxidized one-carbon unit, CO_2, is carried by biotin rather than by tetrahydrofolate.

TABLE 24.2 One-carbon groups carried by tetrahydrofolate

	Group	
Oxidation state	Formula	Name
Most reduced (= methanol)	$—CH_3$	Methyl
Intermediate (= formaldehyde)	$—CH_2—$	Methylene
Most oxidized (= formic acid)	—CHO	Formyl
	—CHNH	Formimino
	—CH═	Methenyl

The one-carbon units carried by tetrahydrofolate are interconvertible (Figure 24.10). N^5,N^{10}-*Methylene*tetrahydrofolate can be reduced to N^5-*methyl*tetrahydrofolate or oxidized to N^5,N^{10}-*methenyl*tetrahydrofolate. N^5,N^{10}-*Methenyl*tetrahydrofolate can be converted into N^5-*formimino*tetrahydrofolate or N^{10}-*formyl*tetrahydrofolate, both of which are at the same oxidation level. N^{10}-*Formyl*tetrahydrofolate can also be synthesized from tetrahydrofolate, formate, and ATP. N^5-*Formyl*tetrahydrofolate can be reversibly isomerized to N^{10}-*formyl*tetrahydrofolate or it can be converted into N^5,N^{10}-*methenyl*tetrahydrofolate.

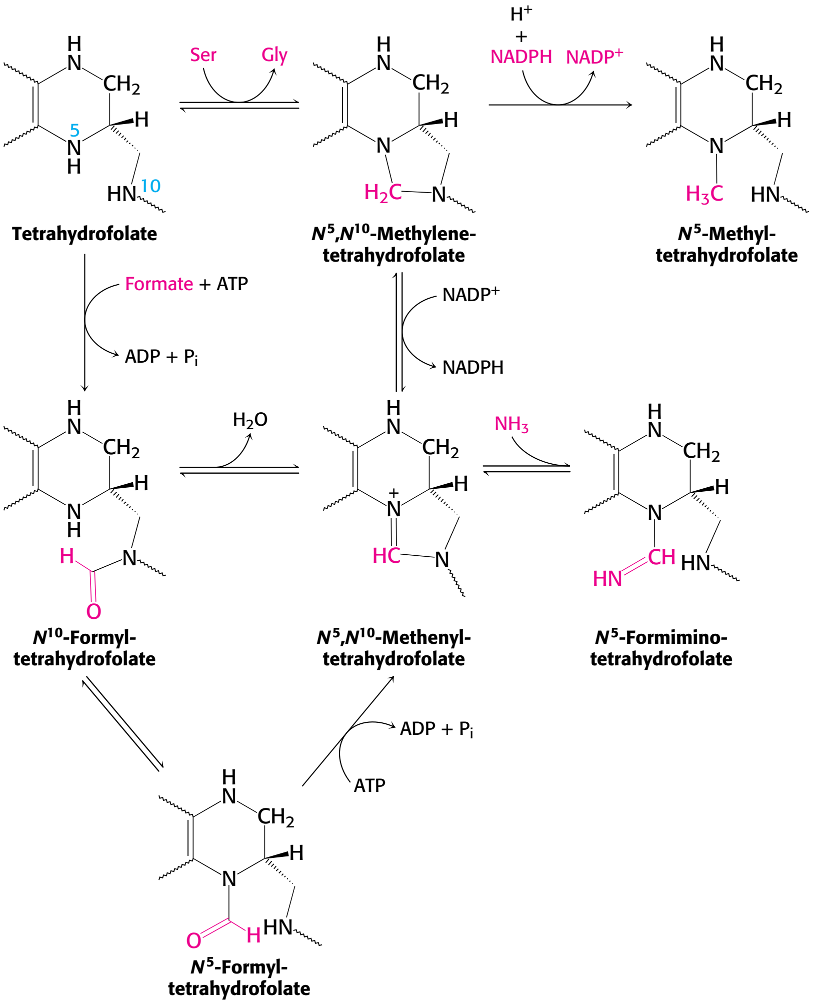

Ser
Gly
H+
+
NADPH
NADP+
Tetrahydrofolate
N5,N10-Methylene-tetrahydrofolate
N5-Methyl-tetrahydrofolate
Formate + ATP
ADP + Pi
NADP+
NADPH
H2O
NH3
N10-Formyl-tetrahydrofolate
N5,N10-Methenyl-tetrahydrofolate
N5-Formimino-tetrahydrofolate
ADP + Pi
ATP
N5-Formyl-tetrahydrofolate

FIGURE 24.10 Conversions of one-carbon units attached to tetrahydrofolate.

These tetrahydrofolate derivatives serve as donors of one-carbon units in a variety of biosyntheses. Methionine is regenerated from homocysteine by transfer of the methyl group of N^5-methyltetrahydrofolate, as will be discussed shortly. We shall see in Chapter 25 that some of the carbon atoms of *purines* are acquired from derivatives of N^{10}-formyltetrahydrofolate. The methyl group of *thymine,* a pyrimidine, comes from N^5, N^{10}-methylenetetrahydrofolate. This tetrahydrofolate derivative can also donate a one-carbon unit in an alternative synthesis of *glycine* that starts with CO_2 and NH_4^+, a reaction catalyzed by *glycine synthase* (called the *glycine cleavage enzyme* when it operates in the reverse direction).

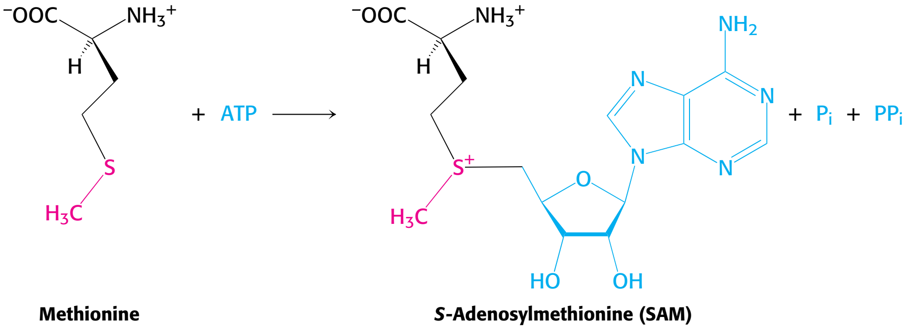
$$CO_2 + NH_4^+ + N^5,N^{10}\text{-methylenetetrahydrofolate} + NADH \rightleftharpoons \text{glycine} + \text{tetrahydrofolate} + NAD^+$$

Thus, one-carbon units at each of the three oxidation levels are utilized in biosyntheses. Furthermore, *tetrahydrofolate serves as an acceptor of one-carbon units in degradative reactions.* The major source of one-carbon units is the facile conversion of serine into glycine by serine hydroxymethyltransferase (p. 723), which yields N^5,N^{10}-methylenetetrahydrofolate. Serine can be derived from 3-phosphoglycerate, and so *this pathway enables one-carbon units to be formed de novo from carbohydrates.*

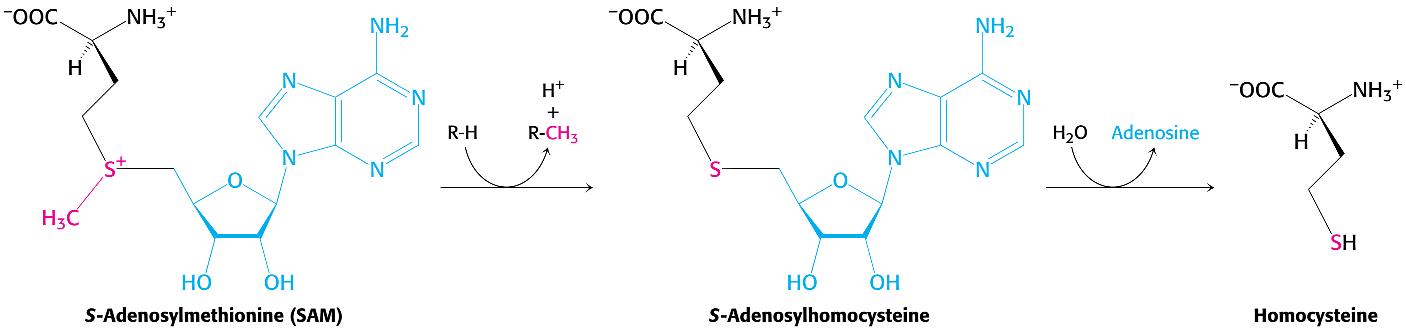
S-Adenosylmethionine is the major donor of methyl groups

Tetrahydrofolate can carry a methyl group on its N-5 atom, but its transfer potential is not sufficiently high for most biosynthetic methylations.

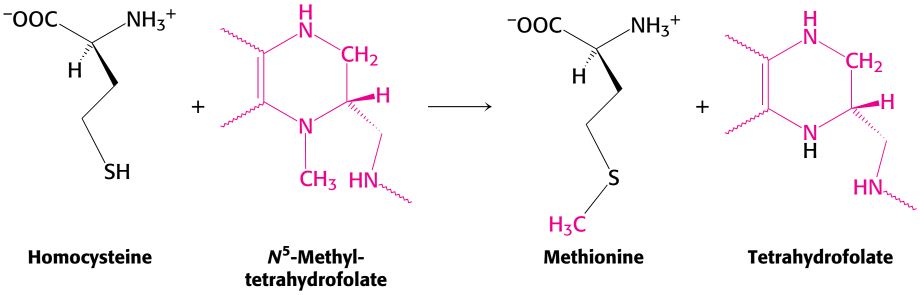
Rather, the activated methyl donor is usually S-*adenosylmethionine* (SAM), which is synthesized by the transfer of an adenosyl group from ATP to the sulfur atom of methionine.


$^{-}$OOC
NH_3^+
H
S
H_3C
+ ATP
NH_2
N
N
N
N
O
S^+
HO
OH
+ P_i + PP_i
Methionine
S-Adenosylmethionine (SAM)

The methyl group of the methionine unit is activated by the positive charge on the adjacent sulfur atom, which makes the molecule much more reactive than N^5-methyltetrahydrofolate. The synthesis of *S*-adenosylmethionine is unusual in that the triphosphate group of ATP is split into pyrophosphate and orthophosphate; the pyrophosphate is subsequently hydrolyzed to two molecules of P_i. S-*Adenosylhomocysteine* is formed when the methyl group of *S*-adenosylmethionine is transferred to an acceptor. *S*-Adenosylhomocysteine is then hydrolyzed to *homocysteine* and adenosine.


$^{-}$OOC
NH_3^+
H
S^+
H_3C
NH_2
N
N
N
N
O
HO
OH
R-H
H^+
+
R-CH_3
S
H_2O
Adenosine
SH
S-Adenosylmethionine (SAM)
S-Adenosylhomocysteine
Homocysteine

Methionine can be regenerated by the transfer of a methyl group to homocysteine from N^5-methyltetrahydrofolate, a reaction catalyzed by *methionine synthase* (also known as *homocysteine methyltransferase*).


$^{-}$OOC
NH_3^+
H
SH
+
H
N
CH_2
H
N
CH_3
HN
S
H_3C
Homocysteine
N^5-Methyl-tetrahydrofolate
Methionine
Tetrahydrofolate

The coenzyme that mediates this transfer of a methyl group is *methylcobalamin*, derived from vitamin B_{12}. In fact, this reaction and the rearrangement of L-methylmalonyl CoA to succinyl CoA (p. 654), catalyzed by a homologous enzyme, are the only two B_{12}-dependent reactions known to take place in mammals. Another enzyme that converts homocysteine into methionine without vitamin B_{12} also is present in many organisms.

These reactions constitute the *activated methyl cycle* (Figure 24.11). Methyl groups enter the cycle in the conversion of homocysteine into

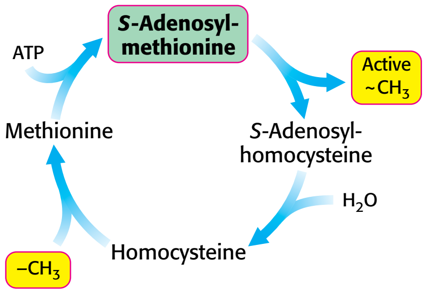
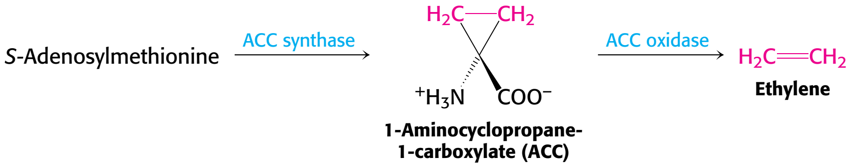

ATP
S-Adenosyl-methionine
Active ~CH_3
Methionine
S-Adenosyl-homocysteine
H_2O
–CH_3
Homocysteine

FIGURE 24.11 Activated methyl cycle. The methyl group of methionine is activated by the formation of *S*-adenosylmethionine.

methionine and are then made highly reactive by the addition of an adenosyl group, which makes the sulfur atoms positively charged and the methyl groups much more electrophilic. The high transfer potential of the *S*-methyl group enables it to be transferred to a wide variety of acceptors. Among the acceptors modified by *S*-adenosylmethionine are specific bases in bacterial DNA. For instance, the methylation of DNA protects bacterial DNA from cleavage by restriction enzymes (Section 9.3). Methylation is also important for the synthesis of phospholipids (Section 26.1).

S-Adenosylmethionine is also the precursor of *ethylene,* a gaseous plant hormone that induces the ripening of fruit. *S*-Adenosylmethionine is cyclized to a cyclopropane derivative that is then oxidized to form ethylene. The Greek philosopher Theophrastus recognized more than 2000 years ago that sycamore figs do not ripen unless they are scraped with an iron claw. The reason is now known: *Wounding triggers ethylene production, which in turn induces ripening.*



Cysteine is synthesized from serine and homocysteine

In addition to being a precursor of methionine in the activated methyl cycle, homocysteine is an intermediate in the synthesis of cysteine. Serine and homocysteine condense to form *cystathionine.* This reaction is catalyzed by *cystathionine β-synthase.* Cystathionine is then deaminated and cleaved to cysteine and α-ketobutyrate by *cystathionine γ-lyase* or *cystathionase*. Both of these enzymes utilize PLP and are homologous to aspartate aminotransferase. The net reaction is

$$\text{Homocysteine} + \text{serine} \rightleftharpoons \text{cysteine} + \alpha\text{-ketobutyrate} + NH_4^+$$

Note that the sulfur atom of cysteine is derived from homocysteine, whereas the carbon skeleton comes from serine.

High homocysteine levels correlate with vascular disease

People with elevated serum levels of homocysteine (homocysteinemia) or the disulfide-linked dimer homocystine (homocystinuria) have an unusually high risk for coronary heart disease and arteriosclerosis. The most common genetic cause of high homocysteine levels is a mutation within the gene encoding cystathionine β-synthase. High levels of homocysteine appear to damage cells lining blood vessels and to increase the growth of vascular smooth muscle. The amino acid raises oxidative stress as well and has also been implicated in the development of type 2 diabetes (Section 27.3). The molecular basis of homocysteine's action has not been clearly identified, but may result from stimulation of the inflammatory response. Vitamin treatments are sometimes effective in reducing homocysteine levels in some people. Treatment with vitamins maximizes the activity of the two major metabolic pathways processing homocysteine. Pyridoxal phosphate, a vitamin B_6 derivative, is necessary for the activity of cystathionine β-synthase, which converts homocysteine into cystathione; tetrahydrofolate, as well as vitamin B_{12}, supports the methylation of homocysteine to methionine.

Shikimate and chorismate are intermediates in the biosynthesis of aromatic amino acids

We turn now to the biosynthesis of essential amino acids. These amino acids are synthesized by plants and microorganisms, and those in the human diet are ultimately derived primarily from plants. The essential amino acids are formed by much more complex routes than are the nonessential amino acids. The pathways for the synthesis of aromatic amino acids in bacteria have been selected for discussion here because they are well understood and exemplify recurring mechanistic motifs.

Phenylalanine, tyrosine, and tryptophan are synthesized by a common pathway in *E. coli* (Figure 24.12). The initial step is the condensation of phosphoenolpyruvate (a glycolytic intermediate) with erythrose 4-phosphate (a pentose phosphate pathway intermediate). The resulting seven-carbon open-chain sugar is oxidized, loses its phosphoryl group, and cyclizes to 3-dehydroquinate. Dehydration then yields 3-dehydroshikimate,

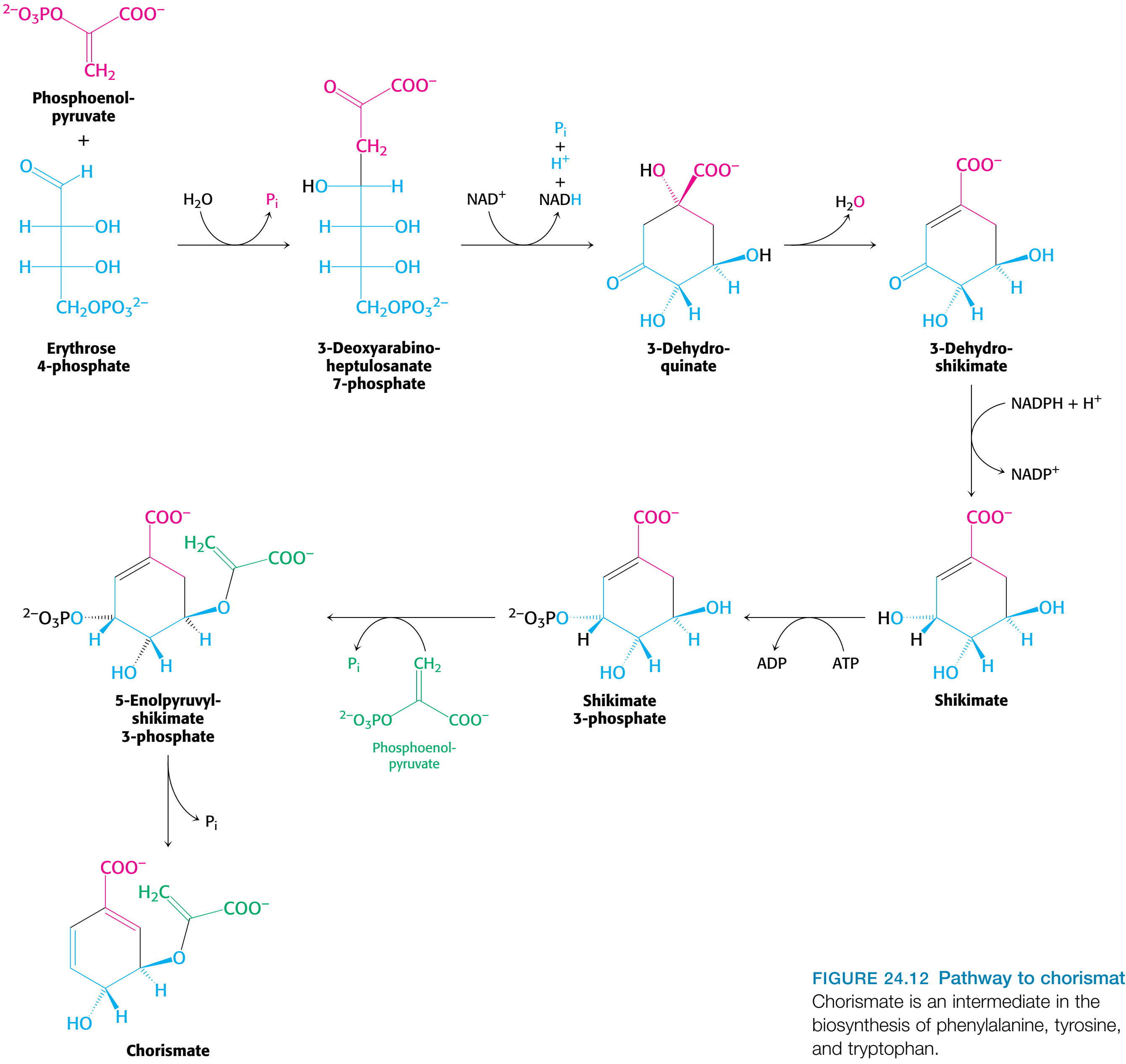

Phosphoenol-pyruvate
+
Erythrose 4-phosphate
H_2O
P_i
3-Deoxyarabino-heptulosanate 7-phosphate
NAD^+
P_i + H^+ + NADH
3-Dehydro-quinate
H_2O
3-Dehydro-shikimate
NADPH + H^+
$NADP^+$
Shikimate
ATP
ADP
Shikimate 3-phosphate
Phosphoenol-pyruvate
P_i
5-Enolpyruvyl-shikimate 3-phosphate
P_i
Chorismate

FIGURE 24.12 Pathway to chorismate. Chorismate is an intermediate in the biosynthesis of phenylalanine, tyrosine, and tryptophan.

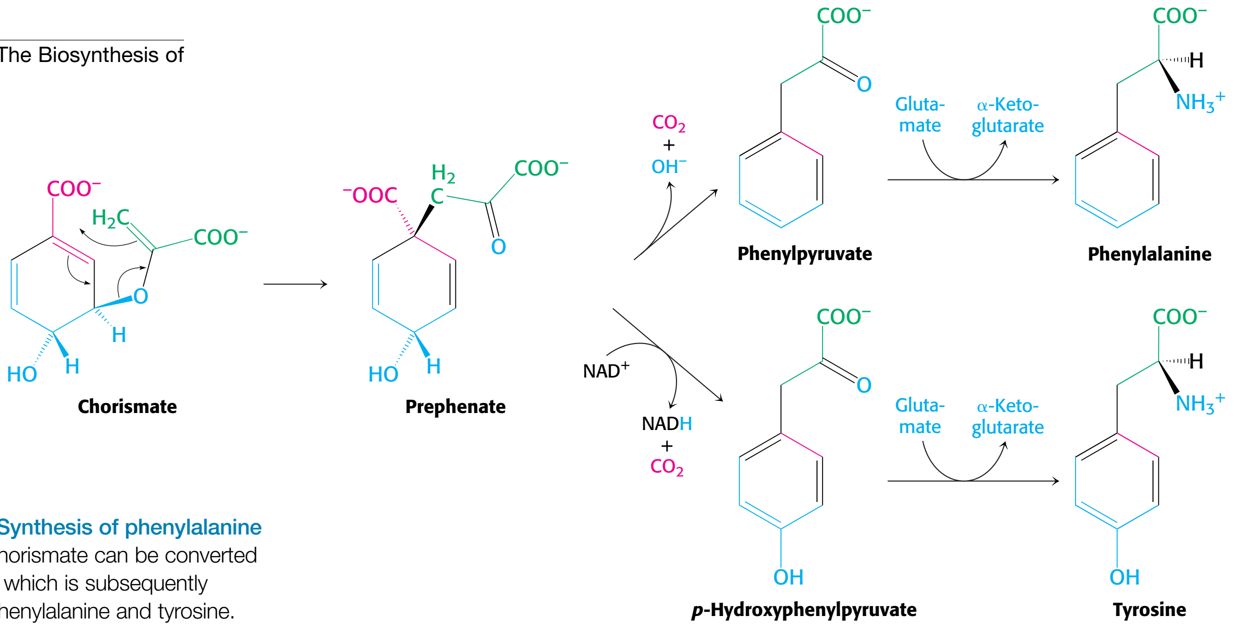
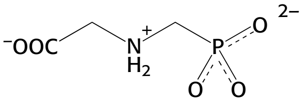


FIGURE 24.13 Synthesis of phenylalanine and tyrosine. Chorismate can be converted into prephenate, which is subsequently converted into phenylalanine and tyrosine.

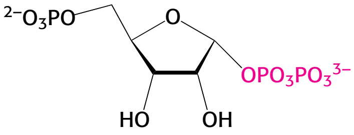
which is reduced by NADPH to shikimate. The phosphorylation of shikimate by ATP gives shikimate 3-phosphate, which condenses with a second molecule of phosphoenolpyruvate. The resulting 5-enolpyruvyl intermediate loses its phosphoryl group, yielding chorismate, the common precursor of all three aromatic amino acids. The importance of this pathway is revealed by the effectiveness of glyphosate (commercially known as Roundup), a broad-spectrum herbicide. This compound is an uncompetitive inhibitor of the enzyme that produces 5-enolpyruvylshikimate 3-phosphate. It blocks aromatic amino acid biosynthesis in plants but is fairly nontoxic in animals because they lack the enzyme.



Glyphosate
(Roundup)

The pathway bifurcates at chorismate. Let us first follow the *prephenate branch* (Figure 24.13). A mutase converts chorismate into prephenate, the immediate precursor of the aromatic ring of phenylalanine and tyrosine. This fascinating conversion is a rare example of an electrocyclic reaction in biochemistry, mechanistically similar to the well-known Diels–Alder reaction in organic chemistry. Dehydration and decarboxylation yield *phenylpyruvate.* Alternatively, prephenate can be oxidatively decarboxylated to p-*hydroxyphenylpyruvate.* These α-ketoacids are then transaminated to form *phenylalanine* and *tyrosine.*

The branch starting with *anthranilate* leads to the synthesis of tryptophan (Figure 24.14). Chorismate acquires an amino group derived from the hydrolysis of the side chain of glutamine and releases pyruvate to form anthranilate. Anthranilate then condenses with *5-phosphoribosyl-1-pyrophosphate (PRPP), an activated form of ribose phosphate.* PRPP is also an important intermediate in the synthesis of histidine, pyrimidine nucleotides, and purine nucleotides (Sections 25.1 and 25.2). The C-1 atom of ribose 5-phosphate becomes bonded to the nitrogen atom of anthranilate in a reaction that is driven by the release and hydrolysis of



5-Phosphoribosyl-1-pyrophosphate (PRPP)

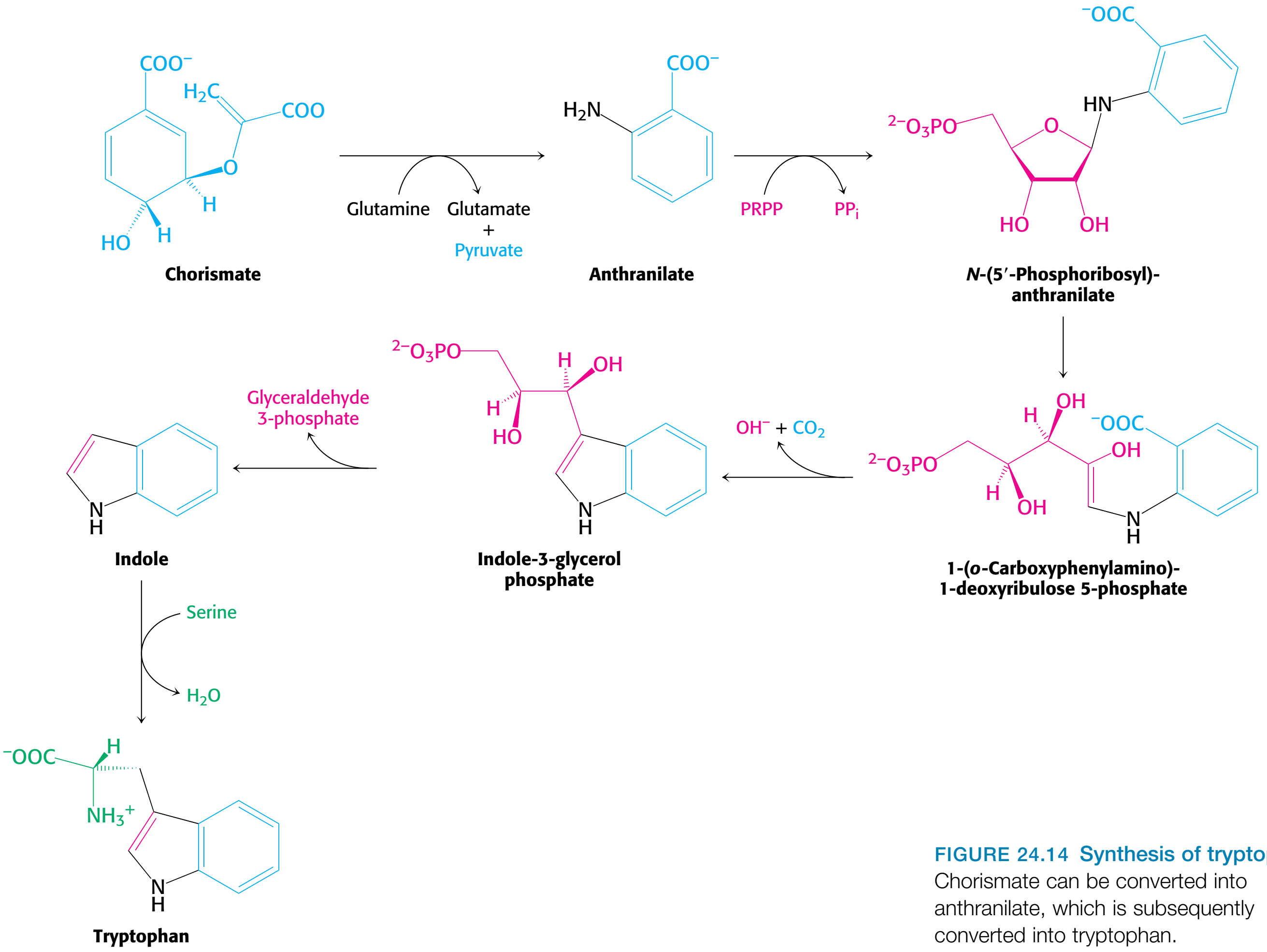

Chorismate
Glutamine
Glutamate
+
Pyruvate
Anthranilate
PRPP
PP$_i$
N-(5′-Phosphoribosyl)-anthranilate
1-(o-Carboxyphenylamino)-1-deoxyribulose 5-phosphate
OH$^-$ + CO$_2$
Indole-3-glycerol phosphate
Glyceraldehyde 3-phosphate
Indole
Serine
H$_2$O
Tryptophan

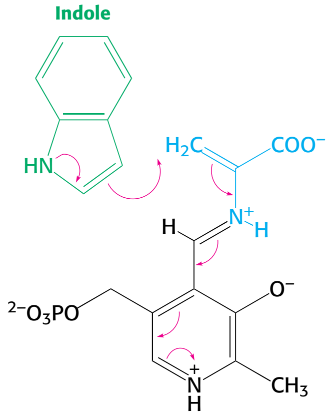
FIGURE 24.14 Synthesis of tryptophan. Chorismate can be converted into anthranilate, which is subsequently converted into tryptophan.

pyrophosphate. The ribose moiety of phosphoribosylanthranilate undergoes rearrangement to yield 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate. This intermediate is dehydrated and then decarboxylated to indole-3-glycerol phosphate. Tryptophan synthase completes the synthesis of tryptophan with the removal of the side chain of indole-3-glycerol phosphate, yielding glyceraldehyde 3-phosphate, and its replacement with the carbon skeleton of serine.

Tryptophan synthase illustrates substrate channeling in enzymatic catalysis

The *E. coli* enzyme *tryptophan synthase,* an $\alpha_2\beta_2$ tetramer, can be dissociated into two α subunits and a β_2 dimer (Figure 24.15). The α subunit catalyzes the formation of indole from indole-3-glycerol phosphate, whereas each β subunit has a PLP-containing active site that catalyzes the condensation of indole and serine to form tryptophan. Serine forms a Schiff base with this PLP, which is then dehydrated to give the *Schiff base of aminoacrylate.* This reactive intermediate is attacked by indole to give tryptophan.

The synthesis of tryptophan poses a challenge. Indole, a hydrophobic molecule, readily traverses membranes and would be lost from the cell if it were allowed to diffuse away from the enzyme. This problem is solved in an ingenious way. A 25-Å-long channel connects the active site of the α subunit with that of the adjacent β subunit in the $\alpha_2\beta_2$ tetramer



Schiff base of aminoacrylate
(derived from serine)

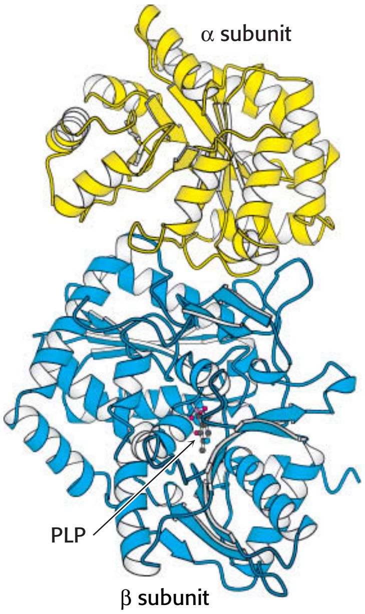


FIGURE 24.15 Structure of tryptophan synthase. The structure of the complex formed by one α subunit (yellow) and one β subunit (blue). *Notice* that pyridoxal phosphate (PLP) is bound deeply inside the β subunit, a considerable distance from the α subunit. [Drawn from 1BKS.pdb.]

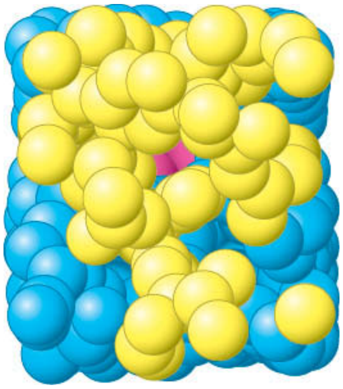


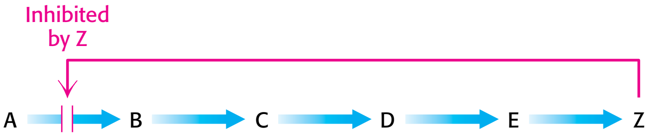
FIGURE 24.16 Substrate channeling. A 25-Å tunnel runs from the active site of the α subunit of tryptophan synthase (yellow) to the PLP cofactor (red) in the active site of the β subunit (blue).

(Figure 24.16). Thus, indole can diffuse from one active site to the other without being released into bulk solvent. Isotopic-labeling experiments showed that indole formed by the α subunit does not leave the enzyme when serine is present. Furthermore, the two partial reactions are coordinated. Indole is not formed by the α subunit until the highly reactive aminoacrylate is ready and waiting in the β subunit. We see here a clear-cut example of *substrate channeling* in catalysis by a multienzyme complex. Channeling substantially increases the catalytic rate. Furthermore, a deleterious side reaction—in this case, the potential loss of an intermediate—is prevented. We shall encounter other examples of substrate channeling in Chapter 25.

24.3 Feedback Inhibition Regulates Amino Acid Biosynthesis

The rate of synthesis of amino acids depends mainly on the *amounts* of the biosynthetic enzymes and on their *activities*. We now consider the control of enzymatic activity. The regulation of enzyme synthesis will be discussed in Chapter 31.

In a biosynthetic pathway, the first irreversible reaction, called the *committed step*, is usually an important regulatory site. The final product of the pathway (Z) often inhibits the enzyme that catalyzes the committed step ($A \rightarrow B$).



This kind of control is essential for the conservation of building blocks and metabolic energy. Consider the biosynthesis of serine (p. 722). The committed step in this pathway is the oxidation of 3-phosphoglycerate, catalyzed by the enzyme *3-phosphoglycerate dehydrogenase.* The *E. coli* enzyme is a tetramer of four identical subunits, each comprising a catalytic domain and a serine-binding regulatory domain (Figure 24.17). The binding of serine

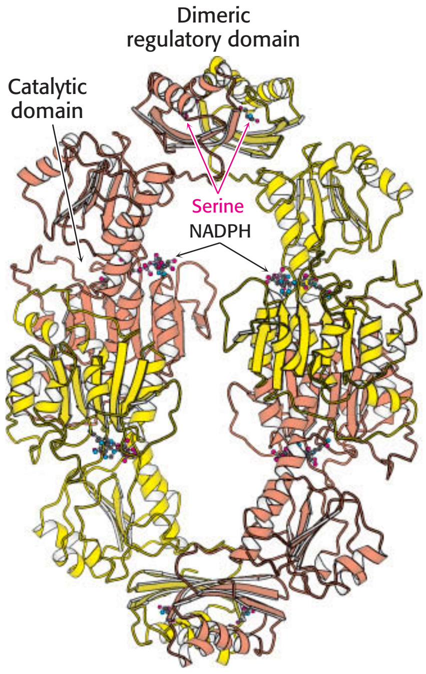


FIGURE 24.17 Structure of 3-phosphoglycerate dehydrogenase. This enzyme, which catalyzes the committed step in the serine biosynthetic pathway, is inhibited by serine. *Notice* the two serine-binding dimeric regulatory domains—one at the top and the other at the bottom of the structure. [Drawn from 1PSD.pdb.]

to a regulatory site reduces the value of V_{max} for the enzyme; an enzyme bound to four molecules of serine is essentially inactive. Thus, if serine is abundant in the cell, the enzyme activity is inhibited, and so 3-phosphoglycerate, a key building block that can be used for other processes, is not wasted.

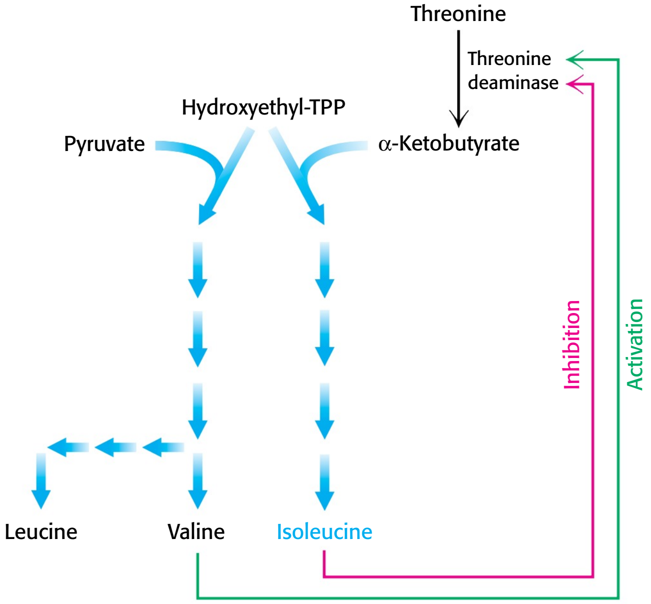


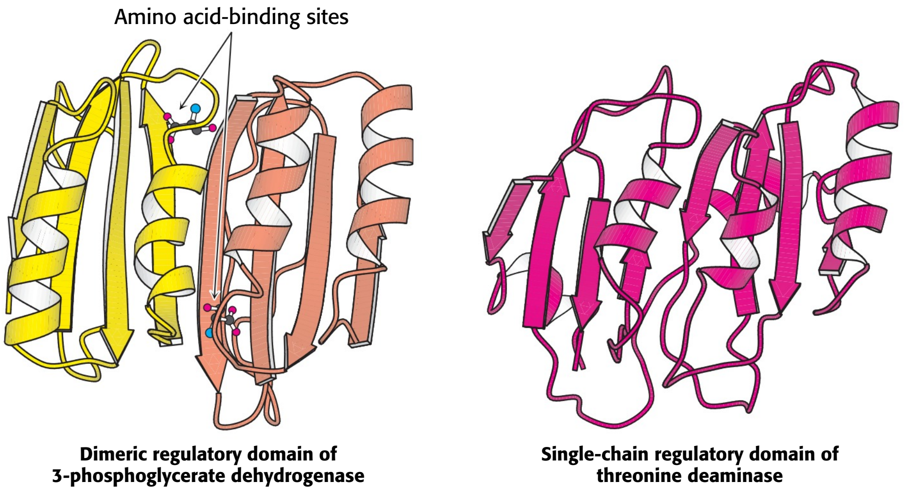
FIGURE 24.18 Regulation of threonine deaminase. Threonine is converted into α-ketobutyrate in the committed step, leading to the synthesis of isoleucine. The enzyme that catalyzes this step, threonine deaminase, is inhibited by isoleucine and activated by valine, the product of a parallel pathway.

Branched pathways require sophisticated regulation

The regulation of branched pathways is more complicated because the concentration of two products must be accounted for. In fact, several intricate feedback mechanisms have been found in branched biosynthetic pathways.

Feedback inhibition and activation. Two pathways with a common initial step may each be inhibited by its own product and activated by the product of the other pathway. Consider, for example, the biosynthesis of the branched chain amino acids valine, leucine, and isoleucine. A common intermediate, hydroxyethyl thiamine pyrophosphate (hydroxyethyl-TPP; Section 17.1), initiates the pathways leading to all three of these amino acids. Hydroxyethyl-TPP reacts with α-ketobutyrate in the initial step for the synthesis of isoleucine. Alternatively, hydroxyethyl-TPP reacts with pyruvate in the committed step for the pathways leading to valine and leucine. Thus, the relative concentrations of α-ketobutyrate and pyruvate determine how much isoleucine is produced compared with valine and leucine. *Threonine deaminase,* the PLP enzyme that catalyzes the formation of α-ketobutyrate, is allosterically inhibited by isoleucine (Figure 24.18). This enzyme is also allosterically activated by valine. Thus, this enzyme is inhibited by the end product of the pathway that it initiates and is activated by the end product of a competitive pathway. This mechanism balances the amounts of different amino acids that are synthesized.

The regulatory domain in threonine deaminase is very similar in structure to the regulatory domain in 3-phosphoglycerate dehydrogenase (Figure 24.19). In the latter enzyme, regulatory domains of two subunits interact to form a dimeric serine-binding regulatory unit, and so the tetrameric enzyme contains two such regulatory units. Each unit is capable of binding two serine molecules. In threonine deaminase, the two regulatory



Dimeric regulatory domain of 3-phosphoglycerate dehydrogenase

Single-chain regulatory domain of threonine deaminase

FIGURE 24.19 A recurring regulatory domain. The regulatory domain formed by two subunits of 3-phosphoglycerate dehydrogenase is structurally related to the single-chain regulatory domain of threonine deaminase. *Notice* that both structures have four α helices and eight β strands in similar locations. Sequence analyses have revealed this amino acid-binding regulatory domain to be present in other enzymes as well. [Drawn from 1PSD and 1TDJ.pdb.]

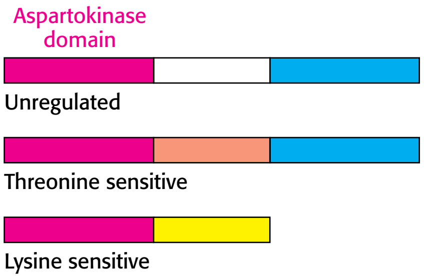
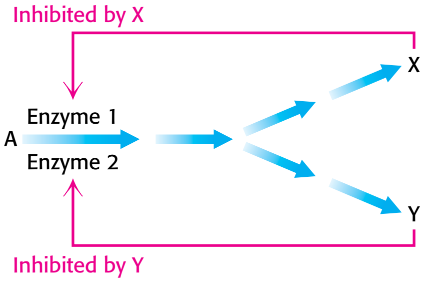


FIGURE 24.20 Domain structures of three aspartokinases. Each catalyzes the committed step in the biosynthesis of a different amino acid: (top) methionine, (middle) threonine, and (bottom) lysine. They have a catalytic domain in common (red) but differ in their regulatory domains (yellow and orange). The blue domain represents another enzyme (homoserine dehydrogenase) involved in aspartate metabolism. Thus, the top two aspartokinases are bifunctional enzymes.

domains are fused into a single unit with two differentiated amino acid-binding sites, one for isoleucine and the other for valine. Sequence analysis shows that similar regulatory domains are present in other amino acid biosynthetic enzymes. *The similarities suggest that feedback-inhibition processes may have evolved by the linkage of specific regulatory domains to the catalytic domains of biosynthetic enzymes.*



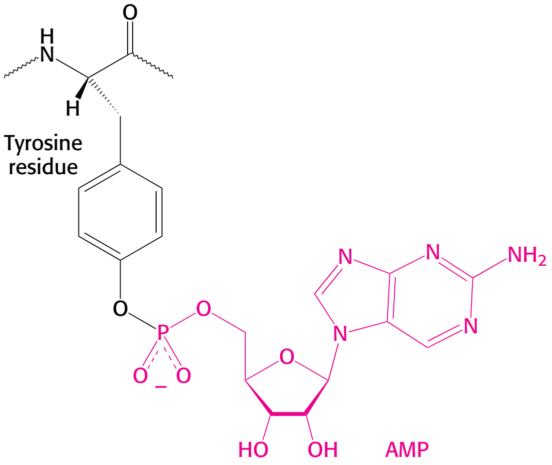
Enzyme multiplicity. The committed step can be catalyzed by two or more isozymes, enzymes with essentially identical catalytic mechanisms, but different regulatory properties (Section 10.2). For example, the phosphorylation of aspartate is the committed step in the biosynthesis of threonine, methionine, and lysine. Three distinct aspartokinases, which evolved by gene duplication, catalyze this reaction in *E. coli* (Figure 24.20). The catalytic domains of these enzymes show approximately 30% sequence identity. Although the mechanisms of catalysis are the same, their activities are regulated differently: one enzyme is not subject to feedback inhibition, another is inhibited by threonine, and the third is inhibited by lysine.

Cumulative feedback inhibition. A common step is partly inhibited by each of the final products, acting independently. The regulation of glutamine synthetase in *E. coli* is a striking example of cumulative feedback inhibition. Recall that glutamine is synthesized from glutamate, NH_4^+, and ATP. *Glutamine synthetase* consists of 12 identical 50-kDa subunits arranged in two hexagonal rings that face each other. This enzyme regulates the flow of nitrogen and hence plays a key role in controlling bacterial metabolism. The amide group of glutamine is a source of nitrogen in the biosyntheses of a variety of compounds, such as tryptophan, histidine, carbamoyl phosphate, glucosamine 6-phosphate, cytidine triphosphate, and adenosine monophosphate. Glutamine synthetase is cumulatively inhibited by each of these final products of glutamine metabolism, as well as by alanine and glycine. *In cumulative inhibition, each inhibitor can reduce the activity of the enzyme, even when other inhibitors are bound at saturating levels.* The enzymatic activity of glutamine synthetase is switched off almost completely when all final products are bound to the enzyme.

The sensitivity of glutamine synthetase to allosteric regulation is altered by covalent modification

The activity of glutamine synthetase is also controlled by *reversible covalent modification*—the attachment of an *AMP unit* by a phosphodiester linkage to the hydroxyl group of a specific tyrosine residue in each subunit. *This adenylylated enzyme is less active and more susceptible to cumulative feedback inhibition than is the deadenylylated form.* The covalently attached AMP unit is removed from the adenylylated enzyme by phosphorolysis.

The adenylylation and phosphorolysis reactions are catalyzed by the same enzyme, *adenylyl transferase.* The adenylyl transferase is composed of two homologous halves, one half catalyzing the adenylylation reaction and the other half the phosphorolytic deadenylylation reaction. What determines whether an AMP unit is added or removed? The specificity of adenylyl transferase is controlled by a *regulatory protein* P_{II}, a trimeric protein that can exist in two forms, unmodified (P_{II}) or covalently bound to UMP (P_{II}-UMP). The complex of P_{II} and adenylyl transferase catalyzes the attachment of an AMP unit to glutamine synthetase, which reduces its activity. Conversely, the complex of P_{II}-UMP and adenylyl transferase removes AMP from the adenylylated enzyme (Figure 24.21).



Tyrosine residue modified by adenylylation

This scheme of regulation immediately raises the question, how is the modification of P_{II} controlled? P_{II} is converted into P_{II}-UMP by the attachment of uridine monophosphate to a specific tyrosine residue (Figure 24.21). This reaction, which is catalyzed by *uridylyl transferase,* is stimulated by ATP and α-ketoglutarate, whereas it is inhibited by glutamine. In turn, the UMP units on P_{II} are removed by hydrolysis, a reaction promoted by glutamine and inhibited by α-ketoglutarate. These opposing catalytic activities are present on a single polypeptide chain and are controlled so that the enzyme does not simultaneously catalyze uridylylation and hydrolysis. In essence, if glutamine is present, the covalent modification system favors adenylylation and inactivation of glutamine synthetase. If glutamine is absent, as indicated by the presence of its precursors α-ketoglutarate and ATP, the control system results in the deadenylylation and activation of the synthetase.

The integration of nitrogen metabolism in a cell requires that a large number of input signals be detected and processed. In addition, the regulatory protein P_{II} also participates in regulating the transcription of genes for glutamine synthetase and other enzymes taking part in nitrogen metabolism. The evolution of covalent regulation superimposed on feedback inhibition provided many more regulatory sites and made possible a finer tuning of the flow of nitrogen in the cell. We have previously seen such a dual regulatory format in the regulation of glycogen metabolism (Section 21.5).

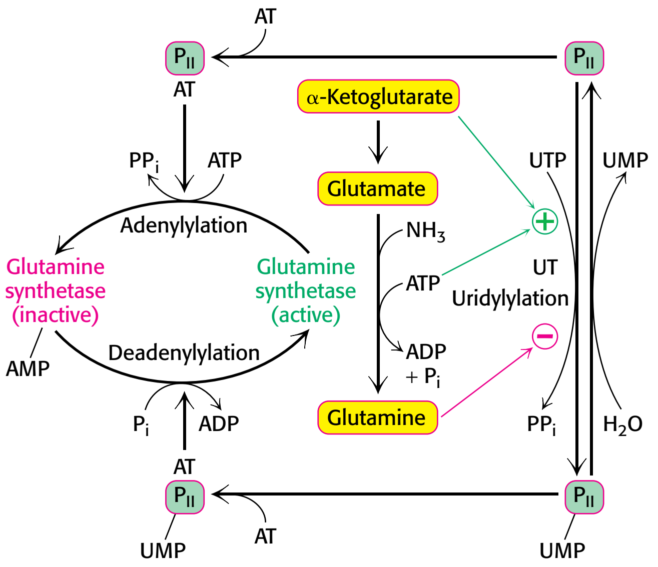


FIGURE 24.21 Covalent regulation of glutamine synthetase. Adenylyl transferase (AT), in association with the regulatory protein P_{II}, adenylylates and inactivates the synthetase. When associated with P_{II} bound to UMP, AT deadenylylates the synthetase, thereby activating the enzyme. Uridylyl transferase (UT), the enzyme that modifies P_{II}, is allosterically regulated by α-ketoglutarate, ATP, and glutamine. [Information from D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry* 6th ed. (W. H. Freeman and Company, 2013), Fig. 22.9.]

24.4 Amino Acids Are Precursors of Many Biomolecules

In addition to being the building blocks of proteins and peptides, amino acids serve as precursors of many kinds of small molecules that have important and diverse biological roles. Let us briefly survey some of the biomolecules that are derived from amino acids (Figure 24.22).

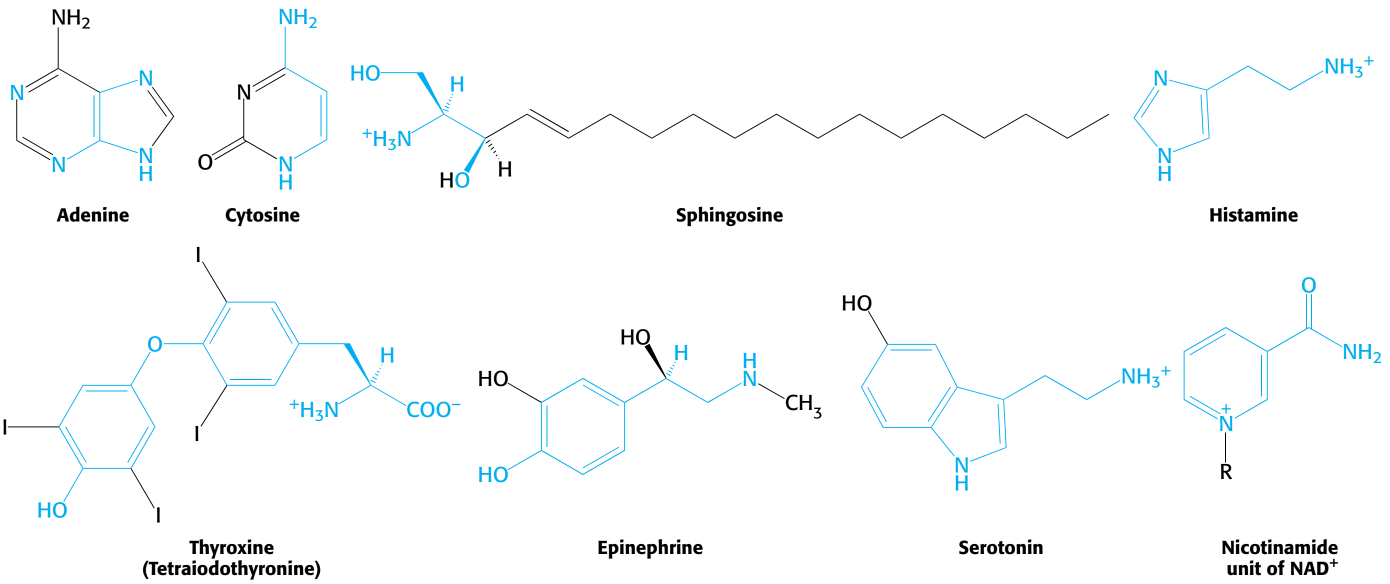


FIGURE 24.22 Selected biomolecules derived from amino acids. The atoms contributed by amino acids are shown in blue.

Purines and *pyrimidines* are derived largely from amino acids. The biosynthesis of these precursors of DNA, RNA, and numerous coenzymes will be discussed in detail in Chapter 25. The reactive terminus of *sphingosine,* an intermediate in the synthesis of sphingolipids, comes from serine. *Histamine,* a potent vasodilator, is derived from histidine by decarboxylation. Tyrosine is a precursor of *thyroxine* (tetraiodothyronine, a hormone that modulates metabolism), *epinephrine* (adrenaline), and *melanin* (a complex polymeric molecule responsible for skin pigmentation). The neurotransmitter *serotonin* (5-hydroxytryptamine) and the *nicotinamide ring* of NAD$^+$ are synthesized from tryptophan. Let us now consider in more detail three particularly important biochemicals derived from amino acids.

Glutathione, a gamma-glutamyl peptide, serves as a sulfhydryl buffer and an antioxidant

Glutathione, a tripeptide containing a sulfhydryl group, is a highly distinctive amino acid derivative with several important roles (Figure 24.23). For example, glutathione, present at high levels (~5 mM) in animal cells, protects red blood cells from oxidative damage by serving as a sulfhydryl buffer (Section 20.5). It cycles between a reduced thiol form (GSH) and an oxidized form (GSSG) in which two tripeptides are linked by a disulfide bond.

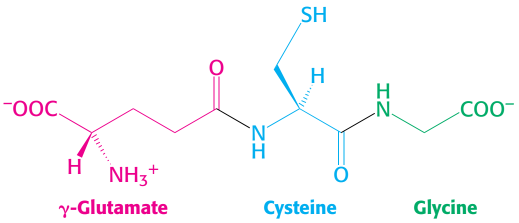


FIGURE 24.23 Glutathione. This tripeptide consists of a cysteine residue flanked by a glycine residue and a glutamate residue that is linked to cysteine by an isopeptide bond between glutamate's side-chain carboxylate group and cysteine's amino group.

$$2\,\text{GSH} + \text{RO—OH} \rightleftharpoons \text{GSSG} + \text{H}_2\text{O} + \text{ROH}$$

GSSG is reduced to GSH by *glutathione reductase,* a flavoprotein that uses NADPH as the electron source. The ratio of GSH to GSSG in most cells is greater than 500. *Glutathione plays a key role in detoxification by reacting with hydrogen peroxide and organic peroxides, the harmful by-products of aerobic life.*

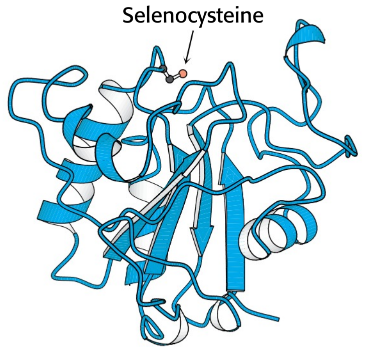


FIGURE 24.24 Structure of glutathione peroxidase. This enzyme, which has a role in peroxide detoxification, contains a selenocysteine residue in its active site. [Drawn from 1GP1.pdb.]

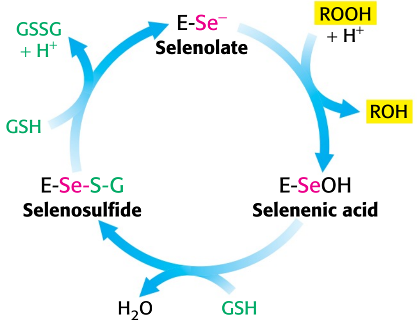


FIGURE 24.25 Catalytic cycle of glutathione peroxidase. [Information from O. Epp, R. Ladenstein, and A. Wendel. *Eur. J. Biochem.* 133(1983):51–69.]

Glutathione peroxidase, the enzyme catalyzing this reaction, is remarkable in having a modified amino acid containing a *selenium* (Se) atom (Figure 24.24). Specifically, its active site contains the selenium analog of cysteine, in which selenium has replaced sulfur. The selenolate (E-Se$^-$) form of this residue reduces the peroxide substrate to an alcohol and is in turn oxidized to selenenic acid (E-SeOH). Glutathione then comes into action by forming a selenosulfide adduct (E-Se-S-G). A second molecule of glutathione then regenerates the active form of the enzyme by attacking the selenosulfide to form oxidized glutathione (Figure 24.25).

Nitric oxide, a short-lived signal molecule, is formed from arginine

Nitric oxide (NO) is an important messenger in many vertebrate signal-transduction processes, identified first as a relaxing factor in the cardiovascular system. It is now known to have a variety of roles not only in the cardiovascular system, but also in the immune and nervous systems. NO has also been shown to stimulate mitochondrial biogenesis. This free-radical gas is produced endogenously from *arginine* in a complex reaction that is catalyzed by *nitric oxide synthase.* NADPH and O_2 are required for the synthesis of nitric oxide (Figure 24.26). Nitric oxide acts by binding to and activating soluble guanylate cyclase, an important enzyme in signal transduction (Section 32.3). This enzyme is homologous to adenylate cyclase but includes a heme-containing domain that binds NO.

^{15}N labeling: A pioneer's account

"Myself as a Guinea Pig

. . . in 1944, I undertook, together with David Rittenberg, an investigation on the turnover of blood proteins of man. To this end I synthesized 66 g of glycine labeled with 35 percent ^{15}N at a cost of $1000 for the ^{15}N. On 12 February 1945, I started the ingestion of the labeled glycine. Since we did not know the effect of relatively large doses of the stable isotope of nitrogen and since we believed that the maximum incorporation into the proteins could be achieved by the administration of glycine in some continual manner, I ingested 1 g samples of glycine at hourly intervals for the next 66 hours . . . . At stated intervals, blood was withdrawn and after proper preparation the ^{15}N concentrations of different blood proteins were determined."

—David Shemin
Bioessays 10(1989):30

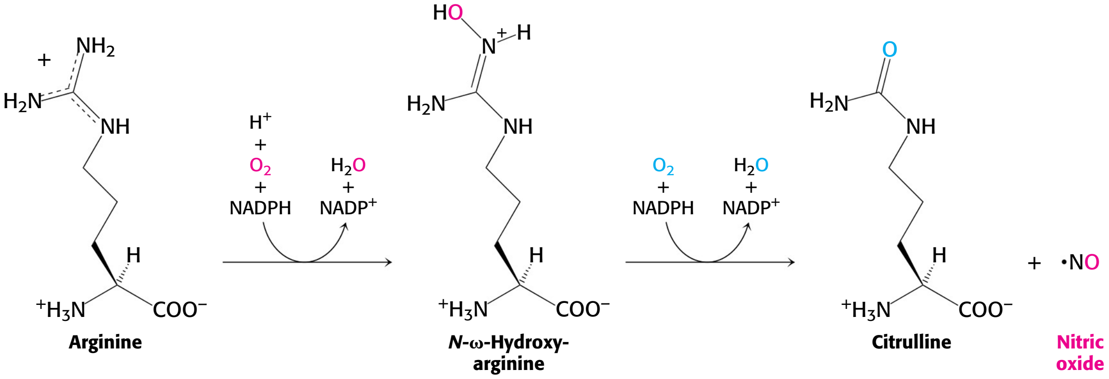


FIGURE 24.26 Formation of nitric oxide. NO is generated by the oxidation of arginine.

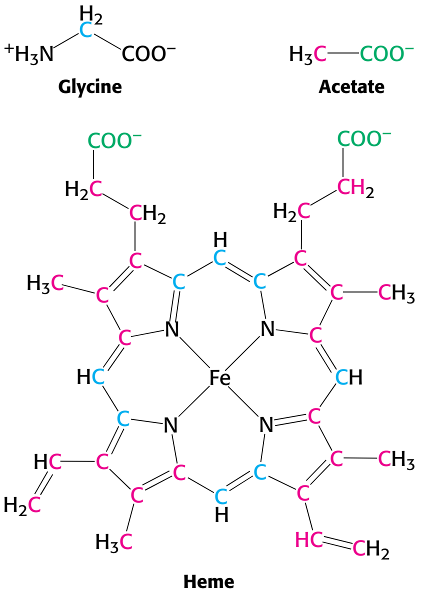


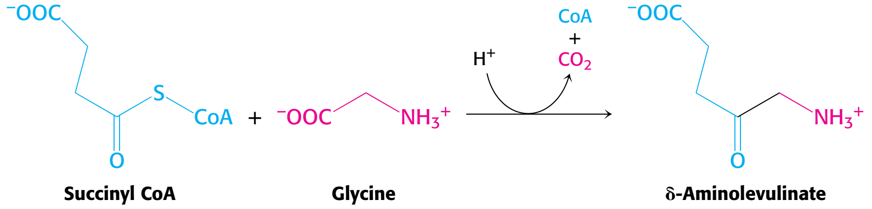
FIGURE 24.27 Heme labeling. The origins of atoms in heme revealed by the results of isotopic labeling studies.

Porphyrins are synthesized from glycine and succinyl coenzyme A

The participation of an amino acid in the biosynthesis of the porphyrin rings of hemes and chlorophylls was first revealed by isotope-labeling experiments carried out by David Shemin and his colleagues. In 1945, they showed that the nitrogen atoms of heme were labeled after the feeding of [^{15}N]glycine to human subjects, whereas the ingestion of [^{15}N]glutamate resulted in very little labeling.

Experiments using ^{14}C, which had just become available, revealed that 8 of the carbon atoms of heme in nucleated duck erythrocytes are derived from the α-carbon atom of glycine and none from the carboxyl carbon atom. Subsequent studies demonstrated that the other 26 carbon atoms of heme can arise from acetate. Moreover, the ^{14}C in methyl-labeled acetate emerged in 24 of these carbon atoms, whereas the ^{14}C in carboxyl-labeled acetate appeared only in the other 2 (Figure 24.27).

This highly distinctive labeling pattern suggested that acetate is converted to succinyl-CoA through enzymes from the citric acid cycle (Section 17.2) and that a heme precursor is formed by the condensation of glycine with succinyl-CoA. Indeed, *the first step in the biosynthesis of porphyrins in mammals is the condensation of glycine and succinyl CoA to form δ-aminolevulinate.*



This reaction is catalyzed by *δ-aminolevulinate synthase,* a PLP enzyme present in mitochondria. Consistent with the labeling studies described earlier, the carbon atom from the carboxyl group of glycine is lost as carbon dioxide, while the α-carbon remains in δ-aminolevulinate.

Two molecules of δ-aminolevulinate condense to form *porphobilinogen,* the next intermediate. Four molecules of porphobilinogen then condense head to tail to form a linear *tetrapyrrole* in a reaction catalyzed by *porphobilinogen deaminase.* The enzyme-bound linear tetrapyrrole then cyclizes to form *uroporphyrinogen III*, which has an asymmetric arrangement of side chains. This reaction requires a *cosynthase.* In the presence of synthase alone, uroporphyrinogen I, the nonphysiological symmetric isomer, is produced. Uroporphyrinogen III is also a key intermediate in the synthesis of vitamin B_{12} by bacteria and that of chlorophyll by bacteria and plants (Figure 24.28).

The porphyrin skeleton is now formed. Subsequent reactions alter the side chains and the degree of saturation of the porphyrin ring (Figure 24.28). *Coproporphyrinogen III* is formed by the decarboxylation of the acetate side chains. The desaturation of the porphyrin ring and the conversion of two of the propionate side chains into vinyl groups yield *protoporphyrin IX*. The chelation of iron finally gives *heme,* the prosthetic group of proteins such as myoglobin, hemoglobin, catalase, peroxidase, and cytochrome *c.* The insertion of the *ferrous* form of iron is catalyzed by *ferrochelatase.* Iron is transported in the plasma by *transferrin,* a protein that binds two ferric ions, and is stored in tissues inside molecules of *ferritin* (Section 32.4).

Studies with ^{15}N-labeled glycine revealed that the normal human erythrocyte has a life span of about 120 days. The first step in the degradation of the heme group is the cleavage of its α-methine bridge to form the green pigment *biliverdin,* a linear tetrapyrrole. The central methine bridge

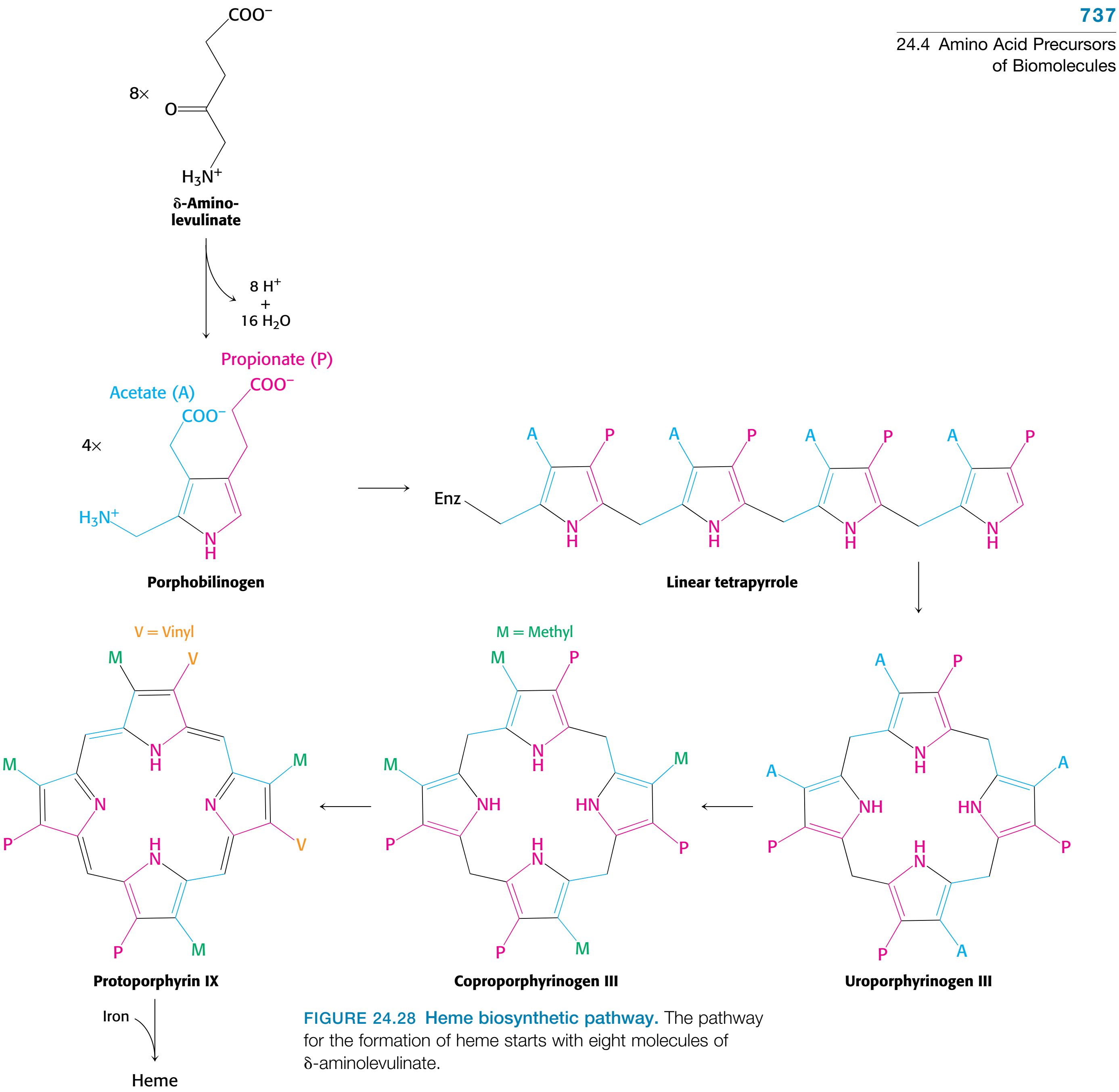

8×
COO⁻
O
H₃N⁺
δ-Amino-levulinate
8 H⁺ + 16 H₂O
Propionate (P)
Acetate (A)
4×
COO⁻
COO⁻
H₃N⁺
N H
Porphobilinogen
Enz
A
P
Linear tetrapyrrole
Uroporphyrinogen III
M = Methyl
Coproporphyrinogen III
V = Vinyl
M
V
Protoporphyrin IX
Iron
Heme

FIGURE 24.28 Heme biosynthetic pathway. The pathway for the formation of heme starts with eight molecules of δ-aminolevulinate.

of biliverdin is then reduced by *biliverdin reductase* to form *bilirubin*, a red pigment (Figure 24.29). The changing color of a bruise is a highly graphic indicator of these degradative reactions.

Porphyrins accumulate in some inherited disorders of porphyrin metabolism

Porphyrias are inherited or acquired disorders caused by a deficiency of enzymes in the heme biosynthetic pathway. Porphyrin is synthesized in both the erythroblasts and the liver, and either one may be the site of a disorder. *Congenital erythropoietic porphyria,* for example, prematurely destroys erythrocytes. This disease results from insufficient cosynthase. In this porphyria, the synthesis of the required amount of uroporphyrinogen III is accompanied by the formation of very large quantities of uroporphyrinogen I, the useless symmetric isomer. Uroporphyrin

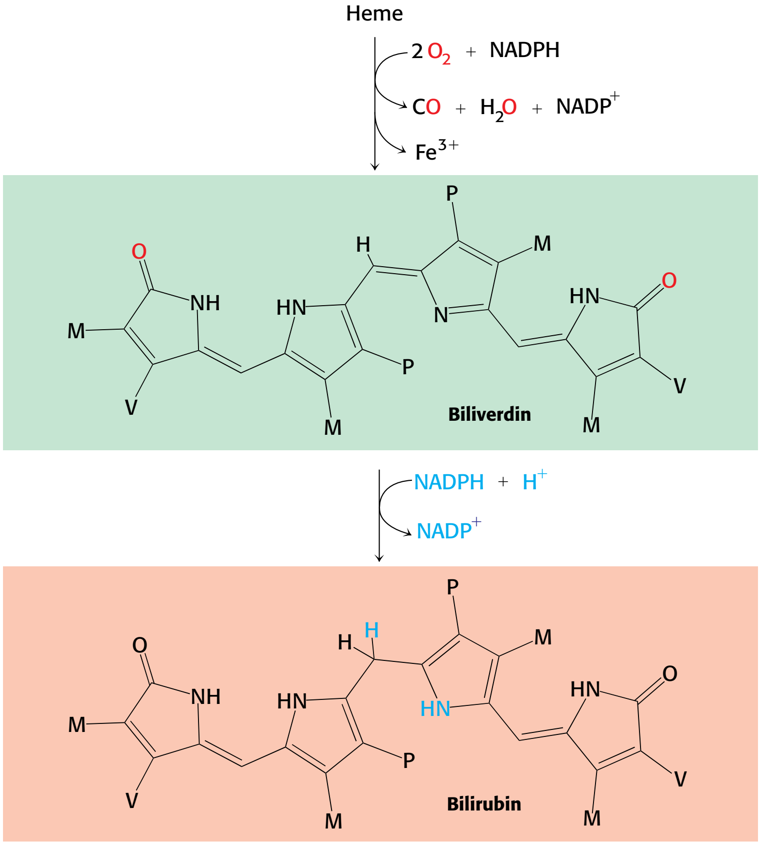


FIGURE 24.29 Heme degradation. The formation of the heme-degradation products biliverdin and bilirubin is responsible for the color of bruises. Abbreviations: M, methyl; V, vinyl.

I, coproporphyrin I, and other symmetric derivatives also accumulate. The urine of patients having this disease is red because of the excretion of large amounts of uroporphyrin I. Their teeth exhibit a strong red fluorescence under ultraviolet light because of the deposition of porphyrins. Furthermore, their skin is usually very sensitive to light because photoexcited porphyrins are quite reactive. *Acute intermittent porphyria* is the most prevalent of the porphyrias affecting the liver. This porphyria is characterized by the overproduction of porphobilinogen and δ-aminolevulinate, which results in severe abdominal pain and neurological dysfunction. The "madness" of George III, king of England during the American Revolution, may have been due to this porphyria.

SUMMARY

24.1 Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia

Microorganisms use ATP and reduced ferredoxin, a powerful reductant, to reduce N_2 to NH_3. An iron–molybdenum cluster in nitrogenase deftly catalyzes the fixation of N_2, a very inert molecule. Higher organisms consume the fixed nitrogen to synthesize amino acids, nucleotides, and other nitrogen-containing biomolecules. The major points of entry of NH_4^+ into metabolism are glutamine or glutamate.

24.2 Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways

Human beings can synthesize 11 of the basic set of 20 amino acids. These amino acids are called nonessential, in contrast with the essential amino acids, which must be supplied in the diet. The pathways for the synthesis of nonessential amino acids are quite simple. Glutamate dehydrogenase catalyzes the reductive amination of α-ketoglutarate to glutamate. A transamination reaction takes place in the synthesis of most amino acids. At this step, the chirality of the amino acid is established. Alanine and aspartate are synthesized by the transamination of pyruvate and oxaloacetate, respectively. Glutamine is synthesized from NH_4^+ and glutamate, and asparagine is synthesized similarly. Proline and arginine are derived from glutamate. Serine, formed from 3-phosphoglycerate, is the precursor of glycine and cysteine. Tyrosine is synthesized by the hydroxylation of phenylalanine, an essential amino acid. The pathways for the biosynthesis of essential amino acids are much more complex than those for the nonessential ones.

Tetrahydrofolate, a carrier of activated one-carbon units, plays an important role in the metabolism of amino acids and nucleotides. This coenzyme carries one-carbon units at three oxidation states, which are interconvertible: most reduced—methyl; intermediate—methylene; and most oxidized—formyl, formimino, and methenyl. The major donor of activated methyl groups is *S*-adenosylmethionine, which is synthesized by the transfer of an adenosyl group from ATP to the sulfur atom of methionine. *S*-Adenosylhomocysteine is formed when the activated methyl group is transferred to an acceptor. It is hydrolyzed to adenosine and homocysteine, and the latter is then methylated to methionine to complete the activated methyl cycle.

24.3 Feedback Inhibition Regulates Amino Acid Biosynthesis

Most of the pathways of amino acid biosynthesis are regulated by feedback inhibition in which the committed step is allosterically inhibited by the final product. The regulation of branched pathways requires extensive interaction among the branches that includes both negative and positive regulation. The regulation of glutamine synthetase in *E. coli* is a striking demonstration of cumulative feedback inhibition and of control by a cascade of reversible covalent modifications.

24.4 Amino Acids Are Precursors of Many Biomolecules

Amino acids are precursors of a variety of biomolecules. Glutathione (γ-Glu-Cys-Gly) serves as a sulfhydryl buffer and detoxifying agent. Glutathione peroxidase, a selenoenzyme, catalyzes the reduction of hydrogen peroxide and organic peroxides by glutathione. Nitric oxide, a short-lived messenger, is formed from arginine. Porphyrins are synthesized from glycine and succinyl CoA, which condense to give δ-aminolevulinate. Two molecules of this intermediate become linked to form porphobilinogen. Four molecules of porphobilinogen combine to form a linear tetrapyrrole, which cyclizes to uroporphyrinogen III. Oxidation and side-chain modifications lead to the synthesis of protoporphyrin IX, which acquires an iron atom to form heme.

KEY TERMS

nitrogen fixation (p. 714)
nitrogenase complex (p. 715)
essential amino acids (p. 719)
nonessential amino acids (p. 719)
pyridoxal phosphate (p. 720)
tetrahydrofolate (p. 723)
S-adenosylmethionine (SAM) (p. 724)
activated methyl cycle (p. 725)
substrate channeling (p. 729)
committed step (p. 730)
enzyme multiplicity (p. 732)
cumulative feedback inhibition (p. 732)
glutathione (p. 734)
nitric oxide (NO) (p. 735)
porphyria (p. 737)

PROBLEMS

1. *Out of thin air.* Define nitrogen fixation. What organisms are capable of nitrogen fixation?

2. *Like Trinidad and Tobago.* Match each term with its description.

(a) Nitrogen fixation ____
(b) Nitrogenase complex ____
(c) Glutamate ____
(d) Essential amino acids ____
(e) Nonessential amino acids ____
(f) Aminotransferase ____
(g) Pyridoxal phosphate ____
(h) Tetrahydrofolate ____
(i) S-Adenosylmethionine ____
(j) Homocysteine ____

1. Methylated to form methionine
2. An important methyl donor
3. Coenzyme required by aminotransferases
4. Conversion of N_2 into NH_3
5. A carrier of various one-carbon units
6. Amino acids that are dietary requirements
7. Amino acids that are readily synthesized
8. Responsible for nitrogen fixation
9. Transfers amino groups between keto acids
10. A common amino group donor

3. *Teamwork.* Identify the two components of the nitrogenase complex and describe their specific tasks.

4. *The fix is in.* "The mechanistic complexity of nitrogenase is necessary because nitrogen fixation is a thermodynamically unfavorable process." True or false? Explain.

5. *Siphoning resources.* Nitrogen-fixing bacteria on the roots of some plants can consume as much as 20% of the ATP produced by their host—consumption that does not seem very beneficial to the plant. Explain why this loss of valuable resources is tolerated and what the bacteria are doing with the ATP.

6. *From few, many.* What are the seven precursors of the 20 amino acids?

7. *Vital, in the truest sense.* Why are certain amino acids defined as essential for human beings?

8. *From sugar to amino acid.* Write a balanced equation for the synthesis of alanine from glucose.

9. *From air to blood.* What are the intermediates in the flow of nitrogen from N_2 to heme?

10. *Common component.* What cofactor is required by all aminotransferases?

11. *Here, hold this.* In this chapter, we considered three different cofactors/cosubstrates that act as carriers of one-carbon units. Name them.

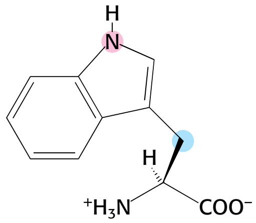
12. *One-carbon transfers.* Which derivative of folate is a reactant in the conversion of (a) glycine into serine? (b) homocysteine into methionine?

13. *Telltale tag.* In the reaction catalyzed by glutamine synthetase, an oxygen atom is transferred from the side chain of glutamate to orthophosphate, as shown by the results of ^{18}O-labeling studies. Account for this finding.

14. *Telltale tag, redux.* In contrast to the production of glutamine by glutamine synthetase (problem 13), the generation of asparagine from ^{18}O-labeled aspartate does not result in the transfer of an ^{18}O atom to orthophosphate. In what molecule do you expect to find one of the ^{18}O atoms?

15. *Therapeutic glycine.* Isovaleric acidemia is an inherited disorder of leucine metabolism caused by a deficiency of isovaleryl CoA dehydrogenase. Many infants having this disease die in the first month of life. The administration of large amounts of glycine sometimes leads to marked clinical improvement. Propose a mechanism for the therapeutic action of glycine.

16. *Lending a hand.* The atoms from tryptophan shaded below are derived from two other amino acids. Name them.



Tryptophan

17. *Deprived bacteria.* Blue-green algae (cyanobacteria) form heterocysts when deprived of ammonia and nitrate. In this form, the cyanobacteria lack nuclei and are attached to adjacent vegetative cells. Heterocysts have photosystem I activity but are entirely devoid of photosystem II activity. What is their role?

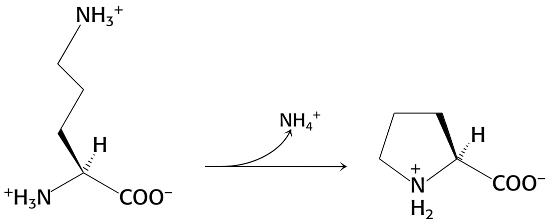
18. *Cysteine and cystine.* Most cytoplasmic proteins lack disulfide bonds, whereas extracellular proteins usually contain them. Why?

19. *Through the looking-glass.* Suppose that aspartate aminotransferase were chemically synthesized with the use of D-amino acids only. What products would you expect if this mirror-image enzyme were treated with (a) L-aspartate and α-ketoglutarate; (b) D-aspartate and α-ketoglutarate?

20. *To and fro.* The synthesis of δ-aminolevulinate takes place in the mitochondrial matrix, whereas the formation of porphobilinogen takes place in the cytoplasm. Propose a reason for the mitochondrial location of the first step in heme synthesis.

21. *Direct synthesis.* Which of the 20 amino acids can be synthesized directly from a common metabolic intermediate by a transamination reaction?

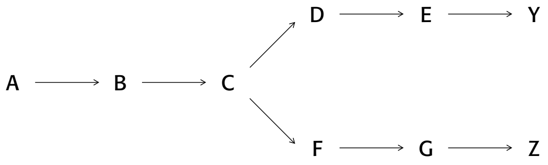
22. *Alternative route to proline.* Certain species of bacteria possess an enzyme, ornithine cyclodeaminase, that can catalyze the conversion of L-ornithine into L-proline in a single catalytic cycle.



Ornithine **Proline**

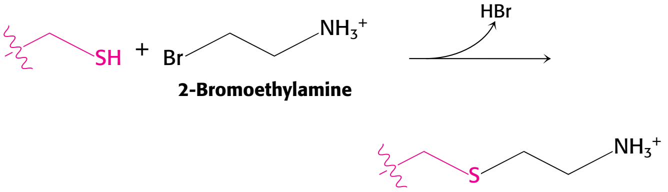
The enzyme *lysine* cyclodeaminase has also been identified. Predict the product of the reaction catalyzed by lysine cyclodeaminase.

23. *Lines of communication.* For the following example of a branched pathway, propose a feedback inhibition scheme that would result in the production of equal amounts of Y and Z.



24. *Cumulative feedback inhibition.* Consider the branched pathway in Problem 23. The first common step (A → B) is partly inhibited by both of the final products, each acting independently of the other. Suppose that a high level of Y alone decreased the rate of the A → B step from 100 to 60 s^{-1} and that a high level of Z alone decreased the rate from 100 to 40 s^{-1}. What would the rate be in the presence of high levels of both Y and Z?

25. *Recovered activity.* Free sulfhydryl groups can be alkylated with 2-bromoethylamine to the corresponding thioether.



Researchers prepared a mutant form of aspartate aminotransferase in which lysine 258 was replaced by cysteine (Lys258Cys). This mutant protein has no observable catalytic activity. However, treatment of Lys258Cys with 2-bromoethylamine yielded a protein with ~7% activity relative to the wild-type enzyme. Explain why alkylation recovered some enzyme activity.

Mechanism Problems

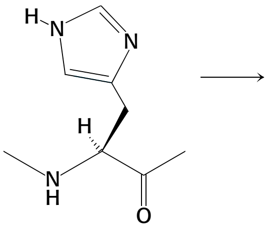
26. *Ethylene formation.* Propose a mechanism for the conversion of *S*-adenosylmethionine into 1-aminocyclopropane-1-carboxylate (ACC) by ACC synthase, a PLP enzyme. What is the other product?

27. *Mirror-image serine.* Brain tissue contains substantial amounts of D-serine, which acts as a neurotransmitter. D-serine is generated from L-serine by serine racemase, a PLP enzyme. Propose a mechanism for the interconversion of L- and D-serine. What is the equilibrium constant for the reaction L-serine $\rightleftharpoons$ D-serine?

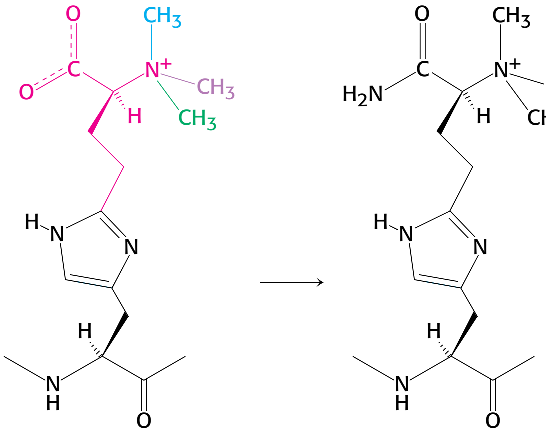
28. *An unusual amino acid.* Elongation factor-2 (eEF-2), a protein taking part in translation, contains a histidine residue that is modified posttranslationally in several steps to a complex side chain known as diphthamide. An intermediate along this pathway is referred to as diphthine.

(a) Labeling experiments indicate that the diphthine intermediate is formed by the modification of histidine with four molecules of *S*-adenosylmethionine (indicated by the four colors). Propose a mechanism for the formation of diphthine.

(b) The final conversion of diphthine into diphthamide is known to be ATP dependent. Propose two possible mechanisms for the final amidation step.



Histidine



Diphthine **Diphthamide**

Chapter Integration Problems

29. *Here, hold this again.* In this chapter, we considered three different cofactors/cosubstrates that act as carriers of one-carbon units (problem 11). Now, name another carrier of one carbon units that we have encountered previously.

30. *Connections.* How might increased synthesis of aspartate and glutamate affect energy production in a cell? How would the cell respond to such an effect?

31. *Protection required.* Suppose that a mutation in bacteria resulted in the diminished activity of methionine adenosyltransferase, the enzyme responsible for the synthesis of SAM from methionine and ATP. Predict how this diminished activity might affect the stability of the mutated bacteria's DNA.

32. *Heme biosynthesis.* Shemin and coworkers used acetate-labeling experiments to conclude that succinyl-CoA is a key intermediate in the biosynthesis of heme. Identify the intermediates in the conversion of acetate into succinyl-CoA.

33. *Comparing K_M.* Glutamate dehydrogenase (p. 717) and glutamine synthetase (p. 717) are present in all organisms. Most prokaryotes also contain another enzyme, *glutamate synthase,* which catalyzes the reductive amination of α-ketoglutarate with the use of glutamine as the nitrogen donor.

$$\alpha\text{-Ketoglutarate} + \text{glutamine} + \text{NADPH} + \text{H}^+ \xrightleftharpoons{\text{Glutamate synthase}} 2\ \text{glutamate} + \text{NADP}^+$$

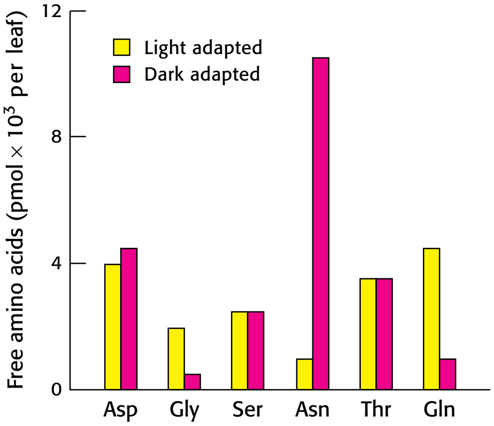
The side-chain amide of glutamine is hydrolyzed to generate ammonia within the enzyme. When NH_4^+ is limiting, most of the glutamate is made by the sequential action of glutamine synthetase and glutamate synthase. The sum of these reactions is

$$\text{NH}_4^+ + \alpha\text{-ketoglutarate} + \text{NADPH} + \text{ATP} \longrightarrow \text{glutamate} + \text{NADP}^+ + \text{ADP} + \text{P}_i$$

Note that this stoichiometry differs from that of the glutamate dehydrogenase reaction in that ATP is hydrolyzed. Why do prokaryotes sometimes use this more-expensive pathway? (Hint: The K_M value for NH_4^+ of glutamate dehydrogenase is higher than that of glutamine synthase.)

Chapter Integration and Data Interpretation Problem

34. *Light effects.* The adjoining graph shows the concentration of several free amino acids in light- and dark-adapted plants.



[Data from B.B. Buchanan, W. Gruissem, and R.L. Jones, *Biochemistry and Molecular Biology of Plants* (American Society of Plant Physiology, 2000), Fig. 8.3, p. 363.]

(a) Of the amino acids shown, which are most affected by light–dark adaptation?

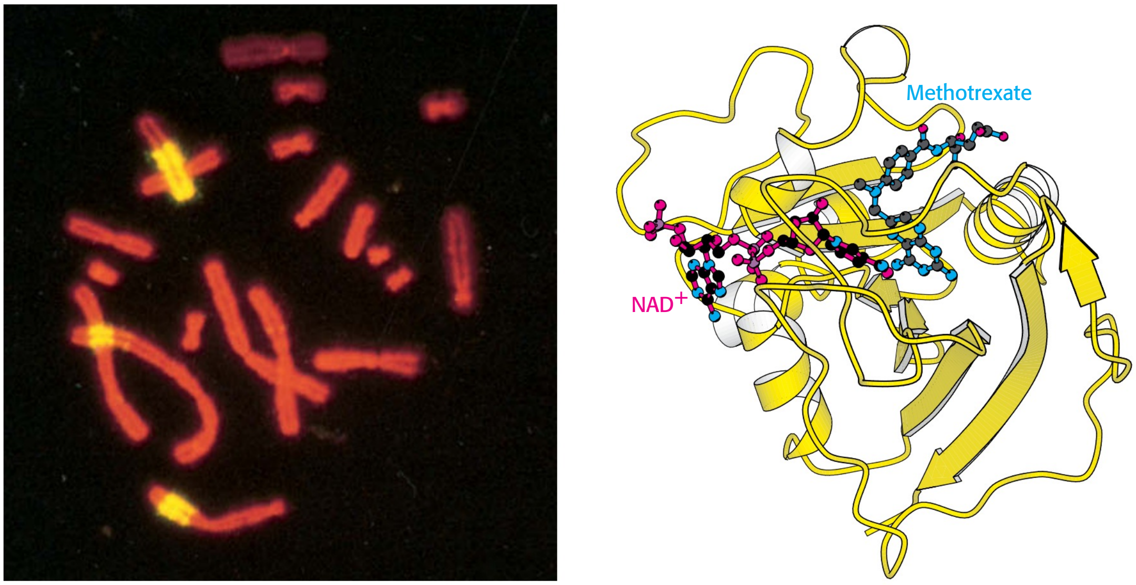
(b) Suggest a plausible biochemical explanation for the difference observed.

(c) White asparagus, a culinary delicacy, is the result of growing asparagus plants in the dark. What chemical might you think enhances the taste of white asparagus?

CHAPTER

25

Nucleotide Biosynthesis



Nucleotides are required for DNA replication and cell division. A key enzyme for the synthesis of one nucleotide is dihydrofolate reductase (right). Cells grown in the presence of methotrexate, a reductase inhibitor, respond by increasing the number of copies of the reductase gene. The bright yellow regions visible on three of the chromosomes in the fluorescence micrograph (left), which were grown in the presence of methotrexate, contain hundreds of copies of the reductase gene. [(Left) Courtesy of Dr. Barbara Trask and Dr. Joyce Hamlin.]

Nucleotides are key biomolecules required for a variety of life processes. First, nucleotides are the *activated precursors of nucleic acids,* necessary for the replication of the genome and the transcription of the genetic information into RNA. Second, an adenine nucleotide, ATP, is *the universal currency of energy.* A guanine nucleotide, GTP, also serves as an energy source for a more select group of biological processes. Third, nucleotide derivatives such as UDP-glucose *participate in biosynthetic processes*, for example, the formation of glycogen. Fourth, nucleotides are *essential components of signal-transduction pathways.* Cyclic nucleotides such as cyclic AMP and cyclic GMP are second messengers that transmit signals both within and between cells. Furthermore, ATP acts as the donor of phosphoryl groups transferred by protein kinases in a variety of signaling pathways and, in some cases, ATP is secreted as a signal molecule.

In this chapter, we continue along the path begun in Chapter 24, which described the incorporation of nitrogen into amino acids from inorganic sources such as nitrogen gas. The amino acids glycine and aspartate are the scaffolds on which the ring systems present in nucleotides are assembled. Furthermore, aspartate and the side chain of glutamine serve as sources of NH_2 groups in the formation of nucleotides.

OUTLINE

25.1 The Pyrimidine Ring Is Assembled de Novo or Recovered by Salvage Pathways

25.2 Purine Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways

25.3 Deoxyribonucleotides Are Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism

25.4 Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition

25.5 Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions

TABLE 25.1 Nomenclature of bases, nucleosides, and nucleotides

RNA

Base	Ribonucleoside	Ribonucleotide (5′-monophosphate)
Adenine (A)	Adenosine	Adenylate (AMP)
Guanine (G)	Guanosine	Guanylate (GMP)
Uracil (U)	Uridine	Uridylate (UMP)
Cytosine (C)	Cytidine	Cytidylate (CMP)

DNA

Base	Deoxyribonucleoside	Deoxyribonucleotide (5′-monophosphate)
Adenine (A)	Deoxyadenosine	Deoxyadenylate (dAMP)
Guanine (G)	Deoxyguanosine	Deoxyguanylate (dGMP)
Thymine (T)	Thymidine	Thymidylate (TMP)
Cytosine (C)	Deoxycytidine	Deoxycytidylate (dCMP)

Nucleotide biosynthetic pathways are tremendously important as intervention points for therapeutic agents. Many of the most widely used drugs in the treatment of cancer block steps in nucleotide biosynthesis, particularly steps in the synthesis of DNA precursors.

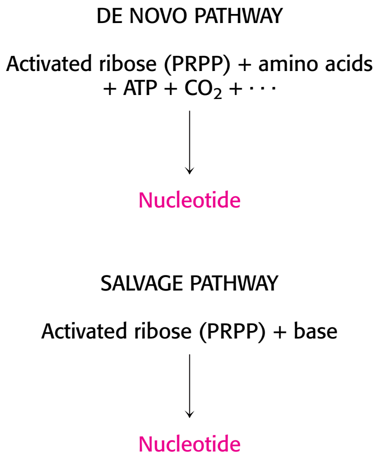

DE NOVO PATHWAY
Activated ribose (PRPP) + amino acids + ATP + CO_2 + · · ·
Nucleotide
SALVAGE PATHWAY
Activated ribose (PRPP) + base
Nucleotide

FIGURE 25.1 De novo and salvage pathways. In de novo synthesis, the base itself is synthesized from simpler starting materials, including amino acids. ATP hydrolysis is required for de novo synthesis. In a salvage pathway, a base is reattached to a ribose, activated in the form of 5-phosphoribosyl-1-pyrophosphate (PRPP).

Nucleotides can be synthesized by de novo or salvage pathways

The pathways for the biosynthesis of nucleotides fall into two classes: *de novo* pathways and *salvage* pathways (Figure 25.1). In de novo (from scratch) pathways, the nucleotide bases are assembled from simpler compounds. The framework for a *pyrimidine* base is assembled first and then attached to ribose. In contrast, the framework for a *purine* base is synthesized piece by piece directly onto a ribose-based structure. These pathways each comprise a small number of elementary reactions that are repeated with variations to generate different nucleotides, as might be expected for pathways that appeared very early in evolution. In salvage pathways, preformed bases are recovered and reconnected to a ribose unit.

De novo pathways lead to the synthesis of *ribo*nucleotides. However, DNA is built from *deoxyribo*nucleotides. Consistent with the notion that RNA preceded DNA in the course of evolution, all deoxyribonucleotides are synthesized from the corresponding ribonucleotides. The deoxyribose sugar is generated by the reduction of ribose within a fully formed nucleotide. Furthermore, the methyl group that distinguishes the thymine of DNA from the uracil of RNA is added at the last step in the pathway.

The nomenclature of nucleotides and their constituent units was presented in Chapter 4. Recall that a *nucleoside* is a purine or pyrimidine base linked to a sugar and that a *nucleotide* is a phosphate ester of a nucleoside. The names of the major bases of RNA and DNA, and of their nucleoside and nucleotide derivatives, are given in Table 25.1.

25.1 The Pyrimidine Ring Is Assembled de Novo or Recovered by Salvage Pathways

In de novo synthesis of pyrimidines, the ring is synthesized first and then it is attached to a ribose phosphate to form a *pyrimidine nucleotide* (Figure 25.2). Pyrimidine rings are assembled from bicarbonate, aspartate, and ammonia. Although an ammonia molecule already present in solution can be used, the ammonia is usually produced from the hydrolysis of the side chain of glutamine.

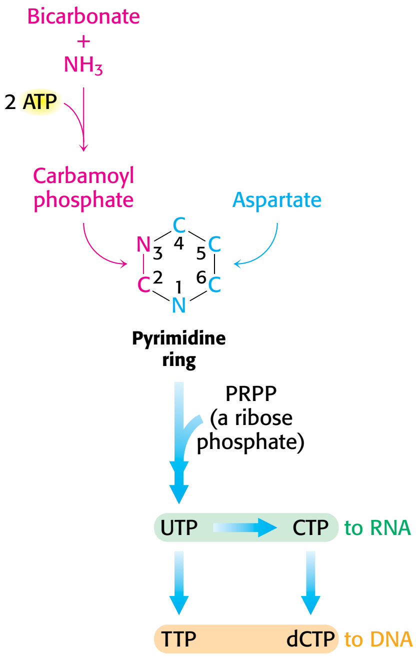


FIGURE 25.2 De novo pathway for pyrimidine nucleotide synthesis. The C-2 and N-3 atoms in the pyrimidine ring come from carbamoyl phosphate, whereas the other atoms of the ring come from aspartate.

Bicarbonate and other oxygenated carbon compounds are activated by phosphorylation

The first step in de novo pyrimidine biosynthesis is the synthesis of *carbamoyl phosphate* from bicarbonate and ammonia in a multistep process, requiring the cleavage of two molecules of ATP. This reaction is catalyzed by *carbamoyl phosphate synthetase II* (CPS II). Recall that carbamoyl phosphate synthetase I facilitates ammonia incorporation into urea (Section 23.4). Carbamoyl phosphate synthetase II is a dimer composed of a small subunit that hydrolyzes glutamine to form NH_3 and a large subunit that completes the synthesis of carbamoyl phosphate. Analysis of the structure of the large subunit reveals two homologous domains, each of which catalyzes an ATP-dependent step (Figure 25.3).

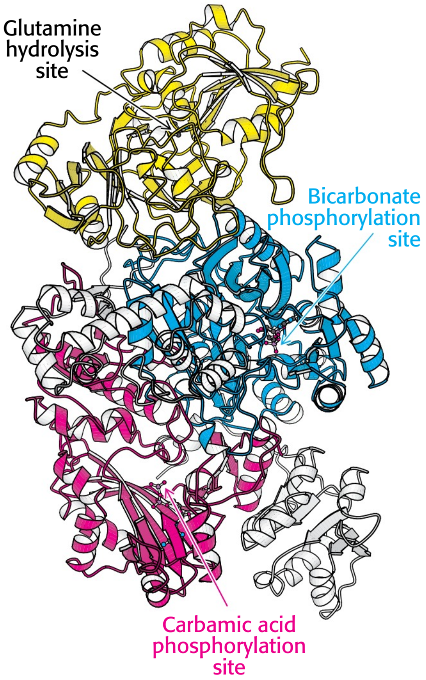
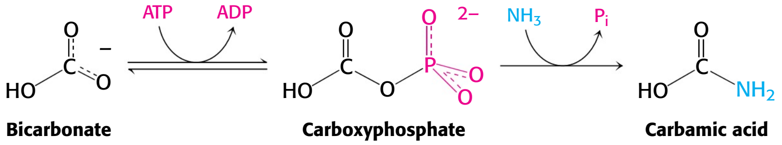

Glutamine hydrolysis site
Bicarbonate phosphorylation site
Carbamic acid phosphorylation site

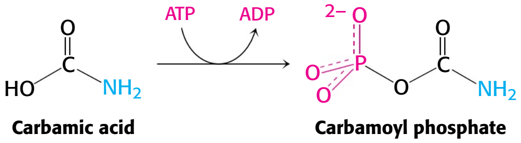
FIGURE 25.3 Structure of carbamoyl phosphate synthetase II. *Notice* that the enzyme contains sites for three reactions. This enzyme consists of two chains. The smaller chain (yellow) contains a site for glutamine hydrolysis to generate ammonia. The larger chain includes two ATP-grasp domains (blue and red). In one ATP-grasp domain (blue), bicarbonate is phosphorylated to carboxyphosphate, which then reacts with ammonia to generate carbamic acid. In the other ATP-grasp domain, the carbamic acid is phosphorylated to produce carbamoyl phosphate. [Drawn from 1JDB.pdb.]

In the first step, bicarbonate is phosphorylated by ATP to form carboxyphosphate and ADP. Ammonia then reacts with carboxyphosphate to form carbamic acid and inorganic phosphate.



The active site for this reaction lies in a domain formed by the amino-terminal third of CPS. This domain forms a structure called an *ATP-grasp fold,* which surrounds ATP and holds it in an orientation suitable for nucleophilic attack at the γ phosphoryl group. Proteins containing ATP-grasp folds catalyze the formation of carbon–nitrogen bonds through acylphosphate intermediates. Such ATP-grasp folds are widely used in nucleotide biosynthesis.

In the second step catalyzed by carbamoyl phosphate synthetase II, carbamic acid is phosphorylated by another molecule of ATP to form carbamoyl phosphate.



This reaction takes place in a second ATP-grasp domain within the enzyme. The active sites leading to carbamic acid formation and carbamoyl phosphate formation are very similar, revealing that this enzyme evolved by a gene duplication event. Indeed, duplication of a gene encoding an ATP-grasp domain followed by specialization was central to the evolution of nucleotide biosynthetic processes (p. 749).

The side chain of glutamine can be hydrolyzed to generate ammonia

Glutamine is the primary source of ammonia for carbamoyl phosphate synthetase II. In this case, the small subunit of the enzyme hydrolyzes glutamine to form ammonia and glutamate. The active site of the glutamine-hydrolyzing component contains a catalytic dyad comprising a cysteine and a histidine residue. Such a catalytic dyad, reminiscent of the active site of cysteine proteases (Figure 9.16), is conserved in a family of amidotransferases, including CTP synthetase and GMP synthetase.

Intermediates can move between active sites by channeling

Carbamoyl phosphate synthetase II contains three different active sites (Figure 25.3), separated from one another by a total of 80 Å. Intermediates

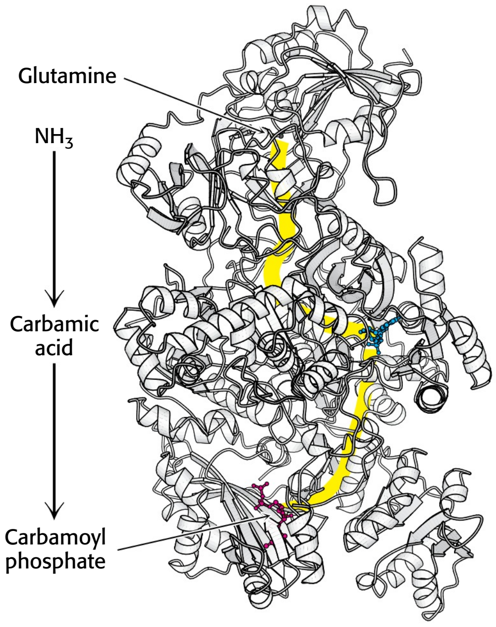
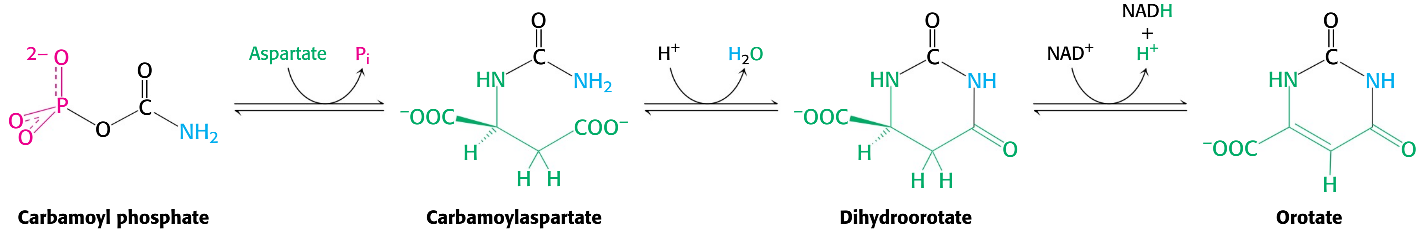


FIGURE 25.4 Substrate channeling. The three active sites of carbamoyl phosphate synthetase II are linked by a channel (yellow) through which intermediates pass. Glutamine enters one active site, and carbamoyl phosphate, which includes the nitrogen atom from the glutamine side chain, leaves another 80 Å away. [Drawn from 1JDB.pdb.]

generated at one site move to the next without leaving the enzyme. These intermediates move within the enzyme by means of substrate channeling, similar to the process described for tryptophan synthetase (Figure 25.4; also Figure 24.16). The ammonia generated in the glutamine-hydrolysis active site travels 45 Å through a channel within the enzyme to reach the site at which carboxyphosphate has been generated. The carbamic acid generated at this site diffuses an additional 35 Å through an extension of the channel to reach the site at which carbamoyl phosphate is generated. This channeling serves two roles: (1) intermediates generated at one active site are captured with no loss caused by diffusion and (2) labile intermediates, such as carboxyphosphate and carbamic acid (which decompose in less than 1 s at pH 7), are protected from hydrolysis. We will see additional examples of substrate channeling later in this chapter.

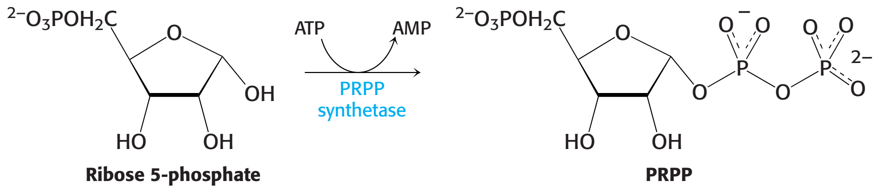
Orotate acquires a ribose ring from PRPP to form a pyrimidine nucleotide and is converted into uridylate

Carbamoyl phosphate reacts with aspartate to form carbamoylaspartate in a reaction catalyzed by *aspartate transcarbamoylase* (Section 10.1). Carbamoylaspartate then cyclizes to form dihydroorotate, which is then oxidized by NAD$^+$ to form orotate.

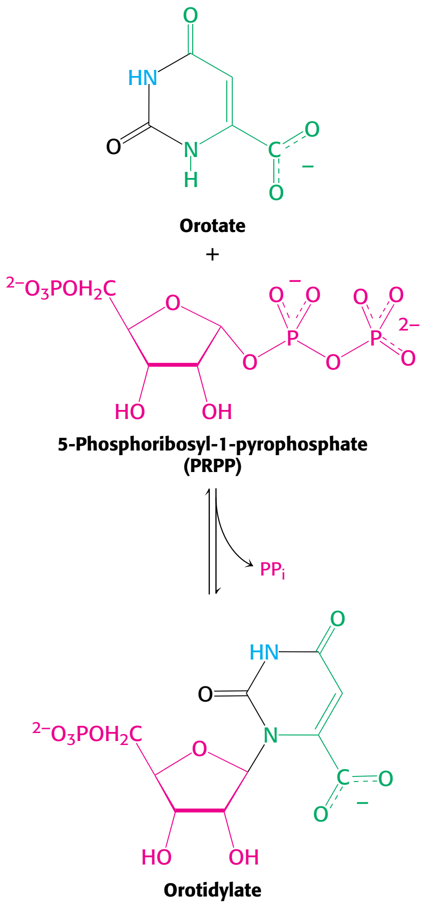


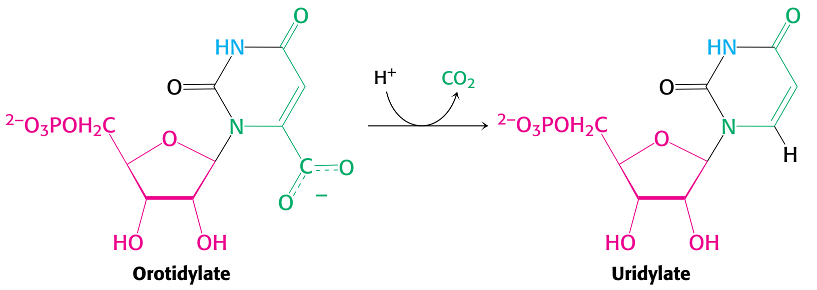
In mammals, the enzymes that form orotate are part of single large polypeptide chain called CAD, for *c*arbamoyl phosphate synthetase, *a*spartate transcarbamoylase and *d*ihydroorotase.

At this stage, orotate couples to ribose, in the form of *5-phosphoribosyl-1-pyrophosphate* (PRPP), a form of ribose activated to accept nucleotide bases. *5-Phosphoribosyl-1-pyrophosphate synthetase* synthesizes PRPP by adding a pyrophosphate from ATP to ribose-5-phosphate, which is formed by the pentose phosphate pathway.



Orotate reacts with PRPP to form *orotidylate,* a pyrimidine nucleotide. This reaction is driven by the hydrolysis of pyrophosphate. The enzyme that catalyzes this addition, *orotate phosphoribosyltransferase,* is homologous to a number of other phosphoribosyltransferases that add different groups to PRPP to form the other nucleotides. Orotidylate is then decarboxylated to form *uridylate* (UMP), a major pyrimidine nucleotide that is a precursor to RNA. This reaction is catalyzed by *orotidylate decarboxylase*, also called orotidine-5′ phosphate decarboxylase.





Orotidylate decarboxylase is one of the most proficient enzymes known. In its absence, decarboxylation is extremely slow and is estimated to take place once every 78 million years; with the enzyme present, it takes place approximately once per second, a rate enhancement of 10^{17}-fold. The phosphoribosyltransferase and decarboxylase activities are located on the same polypeptide chain, providing another example of a bifunctional enzyme. The bifunctional enzyme is called *uridine monophosphate synthetase*.

Nucleotide mono-, di-, and triphosphates are interconvertible

How is the other major pyrimidine ribonucleotide, cytidine, formed? It is synthesized from the uracil base of UMP, but the synthesis can take place only after UMP has been converted into UTP. Recall that the diphosphates and triphosphates are the active forms of nucleotides in biosynthesis and energy conversions. Nucleoside monophosphates are converted into nucleoside triphosphates in stages. First, nucleoside monophosphates are converted into diphosphates by specific *nucleoside monophosphate kinases* that utilize ATP as the phosphoryl-group donor. For example, UMP is phosphorylated to UDP by *UMP kinase.*

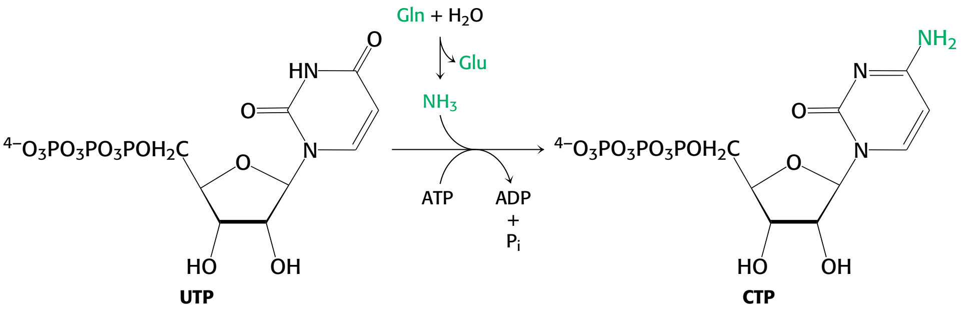
$$\text{UMP} + \text{ATP} \rightleftharpoons \text{UDP} + \text{ADP}$$

Nucleoside diphosphates and triphosphates are interconverted by *nucleoside diphosphate kinase,* an enzyme that has broad specificity, in contrast with the monophosphate kinases. X and Y represent any of several ribonucleosides or even deoxyribonucleosides:

$$\text{XDP} + \text{YTP} \rightleftharpoons \text{XTP} + \text{YDP}$$

CTP is formed by amination of UTP

After uridine triphosphate has been formed, it can be transformed into *cytidine triphosphate* by the replacement of a carbonyl group by an amino group, a reaction catalyzed by *cytidine triphosphate synthetase.*



Like the synthesis of carbamoyl phosphate, this reaction requires ATP and uses glutamine as the source of the amino group. The reaction proceeds through an analogous mechanism in which the O-4 atom is phosphorylated to form a reactive intermediate, and then the phosphate is displaced by ammonia, freed from glutamine by hydrolysis. CTP can then be used in many biochemical processes, including lipid and RNA synthesis.

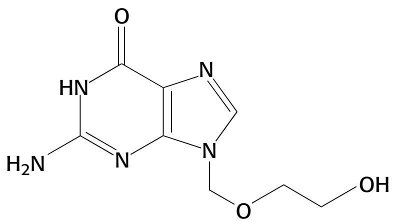
Salvage pathways recycle pyrimidine bases

Pyrimidine bases can be recovered from the breakdown products of DNA and RNA by the use of *salvage pathways*. In these pathways, a preformed base is reincorporated into a nucleotide. We will consider the salvage for the pyrimidine base thymine. Thymine is found in DNA and base-pairs with adenine in the DNA double helix. Thymine released from degraded DNA is salvaged in two steps. First, thymine is converted into the nucleoside thymidine by *thymidine phosphorylase*.

$$\text{Thymine} + \text{deoxyribose-1-phosphate} \rightleftharpoons \text{thymidine} + P_i$$

Thymidine is then converted into a nucleotide by *thymidine kinase*.

$$\text{Thymidine} + \text{ATP} \rightleftharpoons \text{TMP} + \text{ADP}$$



Acyclovir

Viral thymidine kinase differs from the mammalian enzyme and thus provides a therapeutic target. For instance, herpes simplex infections are treated with acyclovir (acycloguanosine), which viral thymidine kinase converts into acyclovir monophosphate by adding a phosphate to the hydroxyl group of acyclovir. Viral thymidine kinase binds acyclovir more than 200 times as tightly as cellular thymidine kinase, accounting for its effect on infected cells only. Acyclovir monophosphate is phosphorylated by cellular enzymes to yield acyclovir triphosphate. Acyclovir triphosphate competes with dGTP for DNA polymerase. Once incorporated into viral DNA, it acts as a chain terminator since it lacks a 3′-hydroxyl required for chain extension. As we will see shortly, thymidine kinase also plays a role in the de novo synthesis of thymidylate.

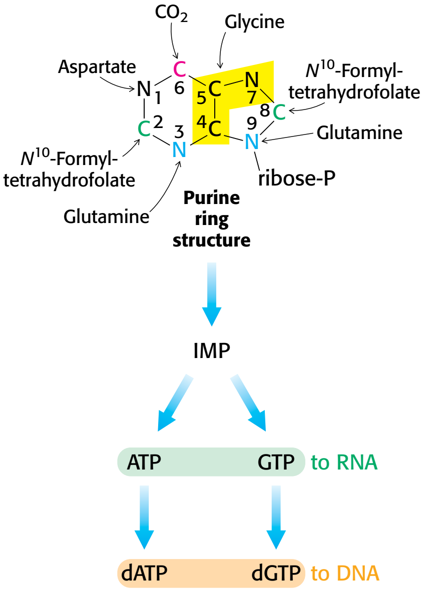


FIGURE 25.5 **De novo pathway for purine nucleotide synthesis.** The origins of the atoms in the purine ring are indicated.

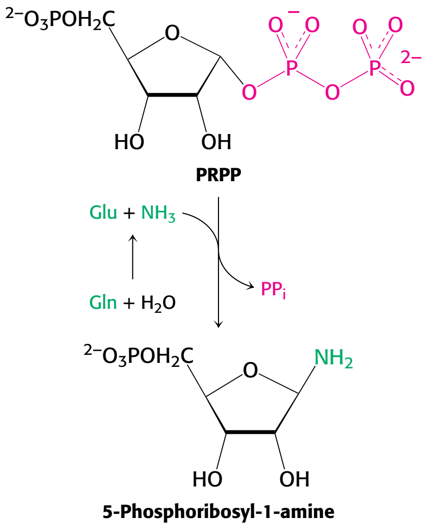
25.2 Purine Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways

Like pyrimidine nucleotides, *purine nucleotides* can be synthesized de novo or by a salvage pathway. When synthesized de novo, purine synthesis begins with simple starting materials such as amino acids and bicarbonate (Figure 25.5). Unlike the bases of pyrimidines, the purine bases are assembled already attached to the ribose ring. Alternatively, purine bases, released by the hydrolytic degradation of nucleic acids and nucleotides, can be salvaged and recycled. Purine salvage pathways are especially noted for the energy that they save and the remarkable effects of their absence (p. 761).

The purine ring system is assembled on ribose phosphate

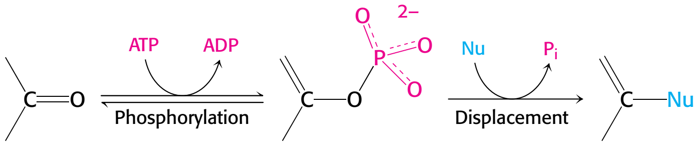
De novo purine biosynthesis, like pyrimidine biosynthesis, requires PRPP but, for purines, PRPP provides the foundation on which the bases are constructed step by step. The initial step is the displacement of pyrophosphate by ammonia, rather than by a preassembled base, to produce *5-phosphoribosyl-1-amine,* with the amine in the β configuration.

Glutamine phosphoribosyl amidotransferase catalyzes this reaction, which is the committed step in purine regulation. This enzyme comprises two domains: the first is homologous to the phosphoribosyltransferases in purine salvage pathways (p. 752), whereas the second produces ammonia from glutamine by hydrolysis. However, this glutamine-hydrolysis domain is distinct from the domain that performs the same function in carbamoyl phosphate synthetase II. In glutamine phosphoribosyl amidotransferase, a cysteine residue located at the amino terminus facilitates glutamine hydrolysis. To prevent wasteful hydrolysis of either substrate, the amidotransferase assumes the active configuration only on binding of both PRPP and glutamine. As is the case with carbamoyl phosphate synthetase II, the ammonia generated at the glutamine-hydrolysis active site passes through a channel to reach PRPP without being released into solution.



The purine ring is assembled by successive steps of activation by phosphorylation followed by displacement

Nine additional steps are required to assemble the purine ring. Remarkably, the first six steps are analogous reactions. Most of these steps are catalyzed by enzymes with ATP-grasp domains that are homologous to those in carbamoyl phosphate synthetase. *Each step consists of the activation of a carbon-bound oxygen atom (typically a carbonyl oxygen atom) by phosphorylation, followed by the displacement of the phosphoryl group by ammonia or an amine group acting as a nucleophile* (Nu).



De novo purine biosynthesis proceeds as shown in Figure 25.6. Table 25.2 lists the enzymes that catalyze each step of the reaction.

1. The carboxylate group of a glycine residue is activated by phosphorylation and then coupled to the amino group of phosphoribosylamine. A new amide bond is formed, and the amino group of glycine is free to act as a nucleophile in the next step.

2. N^{10}-formyltetrahydrofolate donates a formyl moiety to this amino group to form formylglycinamide ribonucleotide.

TABLE 25.2 The enzymes of de novo purine synthesis

Step	Enzyme
1	Glycinamide ribonucleotide (GAR) synthetase
2	GAR transformylase
3	Formylglycinamidine synthase
4	Aminoimidazole ribonucleotide synthetase
5	Carboxyaminoimidazole ribonucleotide synthetase
6	Succinylaminoimidazole carboxamide ribonucleotide synthetase
7	Adenylosuccinate lyase
8	Aminoimidazole carboxamide ribonucleotide transformylase
9	Inosine monophosphate cyclohydrolyase

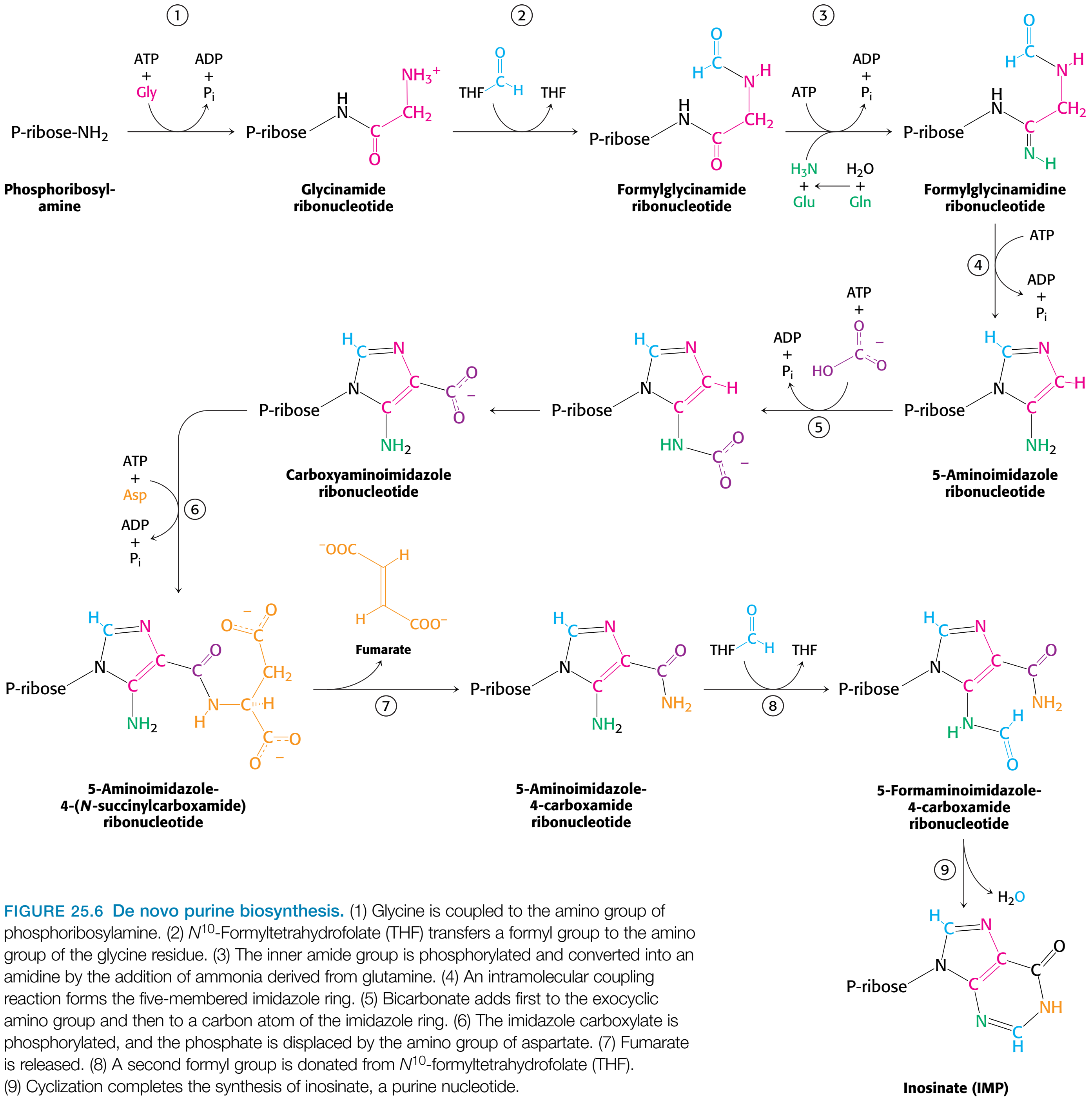

①
ATP + Gly
ADP + P_i
P-ribose-NH_2
Phosphoribosylamine
Glycinamide ribonucleotide
②
THF
THF
Formylglycinamide ribonucleotide
③
ATP
ADP + P_i
H_3N + Glu
H_2O + Gln
Formylglycinamidine ribonucleotide
④
ATP
ADP + P_i
5-Aminoimidazole ribonucleotide
⑤
ATP + HCO_3^-
ADP + P_i
Carboxyaminoimidazole ribonucleotide
⑥
ATP + Asp
ADP + P_i
5-Aminoimidazole-4-(*N*-succinylcarboxamide) ribonucleotide
⑦
Fumarate
5-Aminoimidazole-4-carboxamide ribonucleotide
⑧
THF
THF
5-Formaminoimidazole-4-carboxamide ribonucleotide
⑨
H_2O
Inosinate (IMP)
P-ribose

FIGURE 25.6 De novo purine biosynthesis. (1) Glycine is coupled to the amino group of phosphoribosylamine. (2) N^{10}-Formyltetrahydrofolate (THF) transfers a formyl group to the amino group of the glycine residue. (3) The inner amide group is phosphorylated and converted into an amidine by the addition of ammonia derived from glutamine. (4) An intramolecular coupling reaction forms the five-membered imidazole ring. (5) Bicarbonate adds first to the exocyclic amino group and then to a carbon atom of the imidazole ring. (6) The imidazole carboxylate is phosphorylated, and the phosphate is displaced by the amino group of aspartate. (7) Fumarate is released. (8) A second formyl group is donated from N^{10}-formyltetrahydrofolate (THF). (9) Cyclization completes the synthesis of inosinate, a purine nucleotide.

3. The inner carbonyl group is activated by phosphorylation and then converted into an amidine by the addition of ammonia derived from glutamine.

4. The product of this reaction, formylglycinamidine ribonucleotide, cyclizes to form the five-membered imidazole ring found in purines. Although this cyclization is likely to be favorable thermodynamically, a molecule of ATP is consumed to ensure irreversibility. The familiar pattern is repeated: a phosphoryl group from the ATP molecule activates the carbonyl group and is displaced by the nitrogen atom attached to the ribose molecule. Cyclization is thus an intramolecular reaction in which the nucleophile and phosphate-activated carbon atom are present within the same molecule. In higher eukaryotes, the enzymes catalyzing steps 1, 2, and 4 (Table 25.2) are components of a single polypeptide chain.

5. Bicarbonate is activated by phosphorylation and then attacked by the exocyclic amino group. The product of the reaction in step 5 rearranges to transfer the carboxylate group to the imidazole ring. Interestingly, mammals do not require ATP for this step; bicarbonate apparently attaches directly to the exocyclic amino group and is then transferred to the imidazole ring.

6. The imidazole carboxylate group is phosphorylated again and the phosphate group is displaced by the amino group of aspartate. Once again, in higher eukaryotes, the enzymes catalyzing steps 5 and 6 (Table 25.2) share a single polypeptide chain.

7. Fumarate, an intermediate in the citric acid cycle, is eliminated, leaving the nitrogen atom from aspartate joined to the imidazole ring. The use of aspartate as an amino-group donor and the concomitant release of fumarate are reminiscent of the conversion of citrulline into arginine in the urea cycle, and these steps are catalyzed by homologous enzymes in the two pathways (Section 23.4).

8. A formyl group from N^{10}-formyltetrahydrofolate is added to this nitrogen atom to form a final 5-formaminoimidazole-4-carboxamide ribonucleotide.

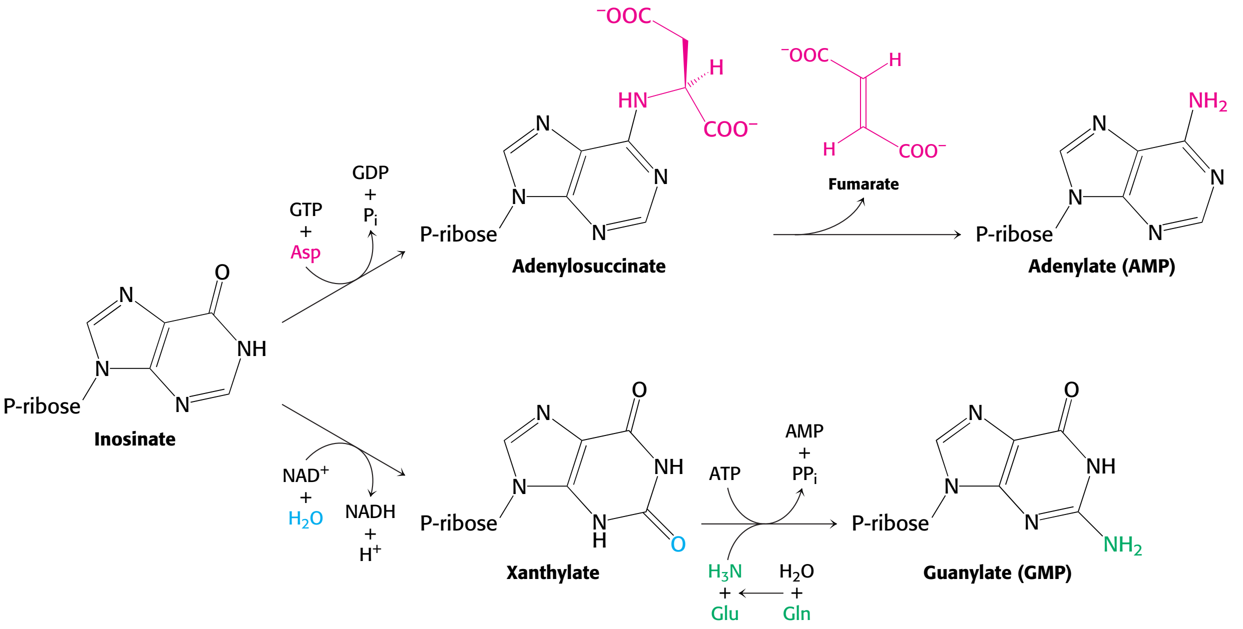
9. 5-Formaminoimidazole-4-carboxamide ribonucleotide cyclizes with the loss of water to form inosinate.

Many of the intermediates in the de novo purine biosynthesis pathway degrade rapidly in water. Their instability in water suggests that the product of one enzyme must be channeled directly to the next enzyme along the pathway. Recent evidence shows that the enzymes do indeed form complexes when purine synthesis is required (p. 752).

AMP and GMP are formed from IMP

A few steps convert inosinate into either AMP or GMP (Figure 25.7). *Adenylate* is synthesized from inosinate by the substitution of an amino group for the carbonyl oxygen atom at C-6. Again, the addition of aspartate followed by the elimination of fumarate contributes the amino group. GTP, rather than ATP, is the phosphoryl-group donor in the synthesis of the adenylosuccinate intermediate from inosinate and aspartate. In accord with the use of GTP, the enzyme that promotes this conversion, *adenylosuccinate synthetase*, is structurally related to the G-protein family and does not contain an ATP-grasp domain.

FIGURE 25.7 Generating AMP and GMP. Inosinate is the precursor of AMP and GMP. AMP is formed by the addition of aspartate followed by the release of fumarate. GMP is generated by the addition of water, dehydrogenation by NAD$^+$, and the replacement of the carbonyl oxygen atom by —NH_2 derived by the hydrolysis of glutamine.



Guanylate is synthesized by the oxidation of inosinate to xanthylate (XMP), followed by the incorporation of an amino group at C-2. NAD^+ is the hydrogen acceptor in the oxidation of inosinate. Xanthylate is activated by the transfer of an AMP group (rather than a phosphoryl group) from ATP to the oxygen atom in the newly formed carbonyl group. Ammonia, generated by the hydrolysis of glutamine, then displaces the AMP group to form guanylate, in a reaction catalyzed by *GMP synthetase*. Note that the synthesis of adenylate requires GTP, whereas the synthesis of guanylate requires ATP. This reciprocal use of nucleotides by the pathways creates an important regulatory opportunity (Section 25.4).

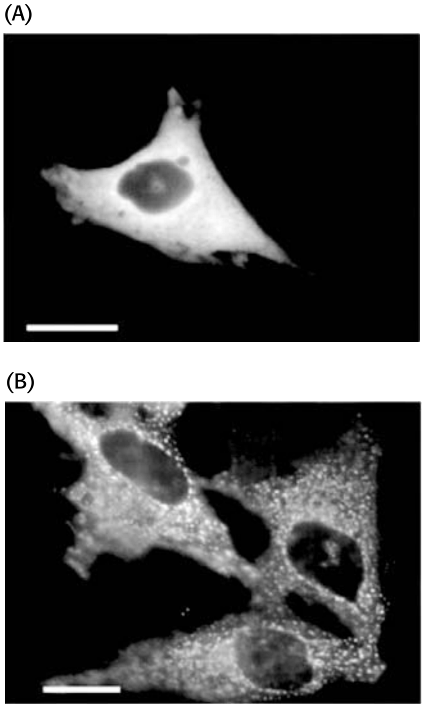


FIGURE 25.8 Formation of purinosomes. A gene construct encoding a fusion protein consisting of formylglycinamidine synthase and GFP was transfected into and expressed in Hela cells, a human cell line. (A) In the presence of purines (the absence of purine synthesis), the GFP was seen as a diffuse stain throughout the cytoplasm. (B) When the cells were shifted to a purine-free medium, purinosomes formed, seen as cytoplasmic granules, and purine synthesis occurred. White scale bars, 10 μm [S. An, R. Kumar, E. D. Sheets, and S. J. Benkovic. *Science* 320:103–106. Figure 2, C and D, 2008.]

Enzymes of the purine synthesis pathway associate with one another in vivo

Biochemists believe that the enzymes of many metabolic pathways, such as glycolysis and the citric acid cycle, are physically associated with one another. Such associations would increase the efficiency of pathways by facilitating the movement of the product of one enzyme to the active site of the next enzyme in the pathway. The evidence for such associations comes primarily from experiments in which one component of a pathway, carefully isolated from the cell, is found to be bound to other components of the pathway. However, these observations raise the question, do enzymes associate with one another in vivo or do they spuriously associate during the isolation procedure? Recent in vivo evidence shows that the enzymes of the purine synthesis pathway associate with one another when purine synthesis is required. Various enzymes of the pathway were fused with the green fluorescent protein (Figure 2.65) and transfected into cells. When cells were grown in the presence of purine, the GFP was spread diffusely throughout the cytoplasm (Figure 25.8A). When the cells were switched to growth media without purines, purine synthesis began and the enzymes became associated with one another, forming complexes dubbed *purinosomes* (Figure 25.8B). The experiments were repeated with other enzymes of the purine synthesis pathway bearing the GFP, and the results were the same: purine synthesis occurs when the enzymes form the purinosomes. What actually causes complex formation? While the results are not yet established, it appears that several G-protein coupled receptors, including those responding to epinephrine as well as ATP and ADP (purinergic receptors), instigate complex formation, while human creatine kinase II (hCK2), responding to the presence of purines, causes disassembly of the purinosome.

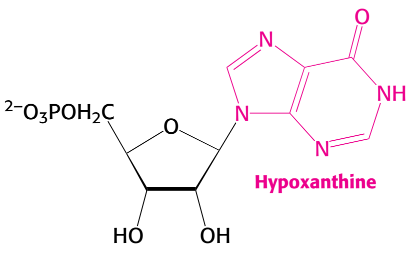
Salvage pathways economize intracellular energy expenditure

As we have seen, the de novo synthesis of purines requires a substantial investment of ATP. Purine salvage pathways provide a more economical means of generating purines. Free purine bases, derived from the turnover of nucleotides or from the diet, can be attached to PRPP to form purine nucleoside monophosphates, in a reaction analogous to the formation of orotidylate. Two salvage enzymes with different specificities recover purine bases. *Adenine phosphoribosyltransferase* catalyzes the formation of adenylate (AMP):

$$\text{Adenine} + \text{PRPP} \longrightarrow \text{adenylate} + PP_i$$

whereas *hypoxanthine-guanine phosphoribosyltransferase* (HGPRT) catalyzes the formation of guanylate (GMP) as well as inosinate (inosine monophosphate, IMP), a precursor of guanylate and adenylate.

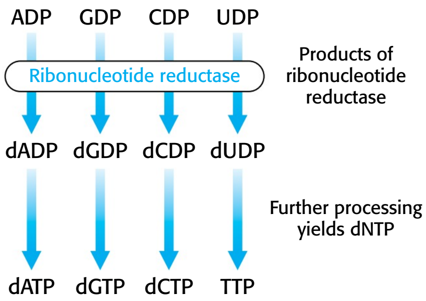
$$\text{Guanine} + \text{PRPP} \longrightarrow \text{guanylate} + PP_i$$
$$\text{Hypoxanthine} + \text{PRPP} \longrightarrow \text{inosinate} + PP_i$$



Inosinate

25.3 Deoxyribonucleotides Are Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism

We turn now to the synthesis of deoxyribonucleotides. These precursors of DNA are formed by the reduction of ribonucleotides; specifically, the 2′-hydroxyl group on the ribose moiety is replaced by a hydrogen atom. The substrates are ribonucleoside diphosphates, and the ultimate reductant is NADPH. The enzyme *ribonucleotide reductase* is responsible for the reduction reaction for all four ribonucleotides. The ribonucleotide reductases of different organisms are a remarkably diverse set of enzymes. Yet detailed studies have revealed that they have a common reaction mechanism, and their three-dimensional structural features indicate that these enzymes are homologous. We will focus on the best understood of these enzymes, that of *E. coli* living aerobically.



Mechanism: A tyrosyl radical is critical to the action of ribonucleotide reductase

The ribonucleotide reductase of *E. coli* consists of two subunits: R1 (an 87-kDa dimer) and R2 (a 43-kDa dimer). The R1 subunit contains the active site as well as two allosteric control sites (Section 25.4). This subunit includes three conserved cysteine residues and a glutamate residue, all four of which participate in the reduction of ribose to deoxyribose (Figure 25.9). The R2 subunit's role in catalysis is to generate a free radical in each of its two chains. Each R2 chain contains a *tyrosyl radical* with an unpaired electron delocalized onto its aromatic ring, generated by a nearby *iron center* consisting of two ferric (Fe^{3+}) ions bridged by an oxide (O^{2-}) ion (Figure 25.10). This very unusual free radical is remarkably stable, with a half-life of 4 days at 4°C. By contrast, free tyrosine radicals in solution have microsecond lifetimes.

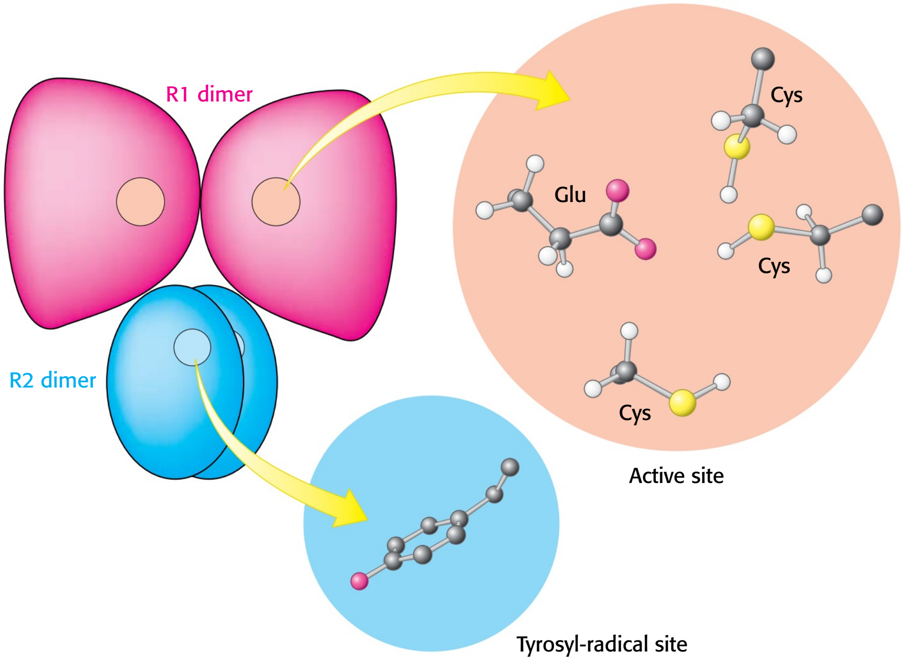


FIGURE 25.9 Ribonucleotide reductase. Ribonucleotide reductase reduces ribonucleotides to deoxyribonucleotides in its active site, which contains three key cysteine residues and one glutamate residue. Each R2 subunit contains a tyrosyl radical that accepts an electron from one of the cysteine residues in the active site to initiate the reduction reaction. Two R1 subunits come together to form a dimer as do two R2 subunits.

In the synthesis of a deoxyribonucleotide, the OH bonded to C-2′ of the ribose ring is replaced by H, with retention of the configuration at the C-2′ carbon atom (Figure 25.11).

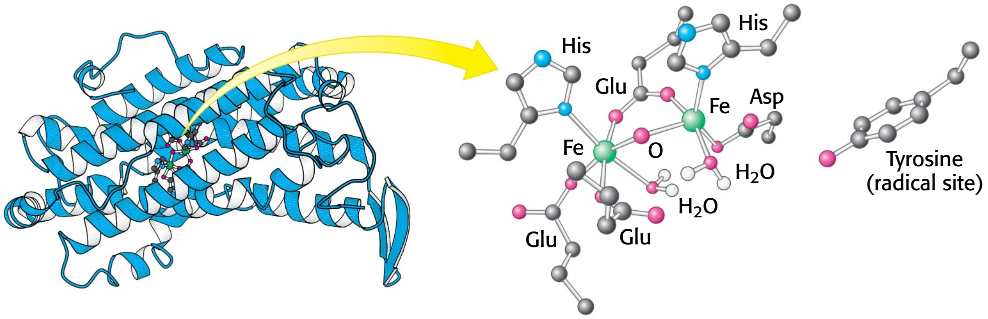


FIGURE 25.10 Ribonucleotide reductase R2 subunit. The R2 subunit contains a stable free radical on a tyrosine residue. This radical is generated by the reaction of oxygen (not shown) at a nearby site containing two iron atoms. Two R2 subunits come together to form a dimer. [Drawn from 1RIB.pdb.]

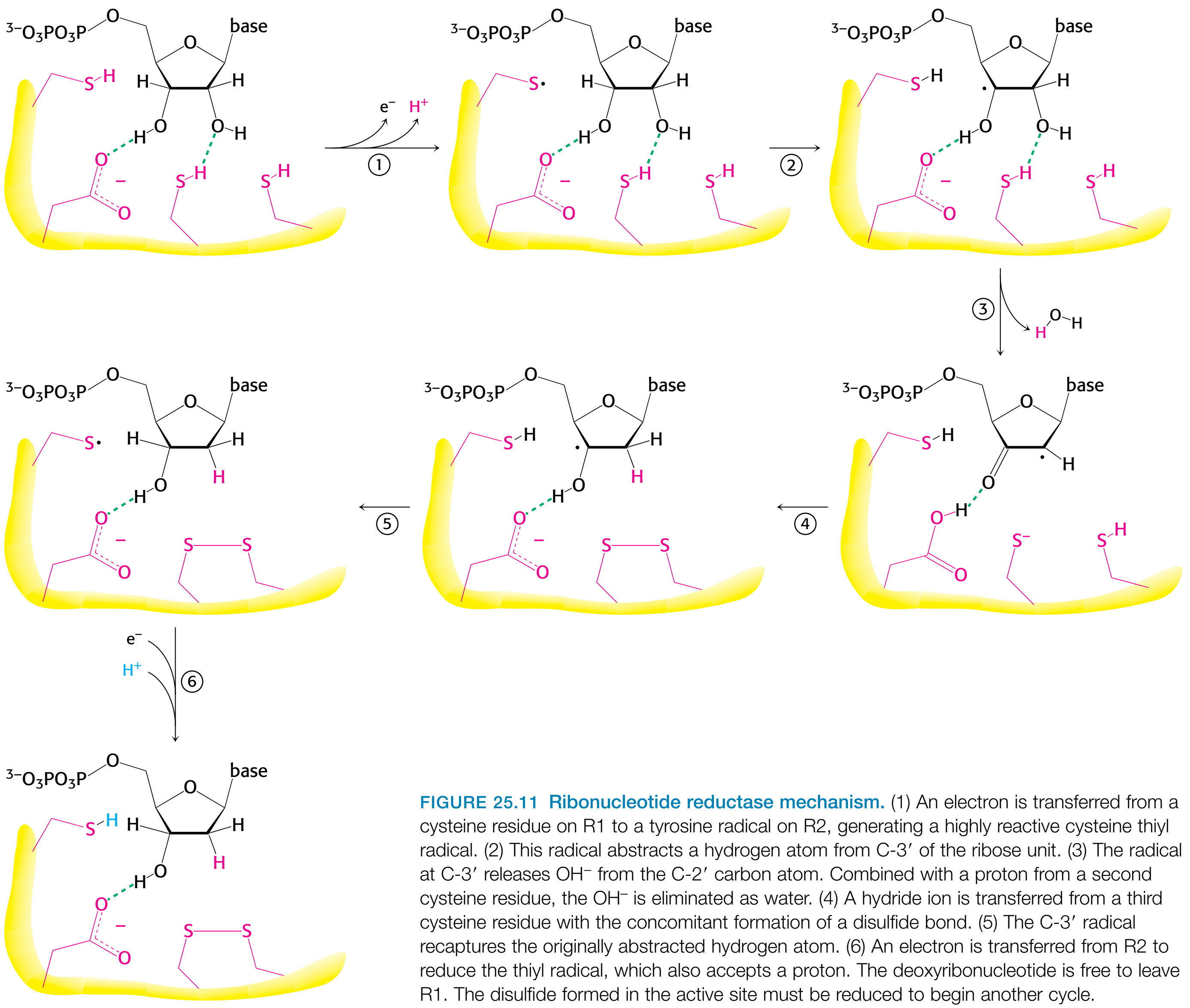


FIGURE 25.11 Ribonucleotide reductase mechanism. (1) An electron is transferred from a cysteine residue on R1 to a tyrosine radical on R2, generating a highly reactive cysteine thiyl radical. (2) This radical abstracts a hydrogen atom from C-3′ of the ribose unit. (3) The radical at C-3′ releases OH^- from the C-2′ carbon atom. Combined with a proton from a second cysteine residue, the OH^- is eliminated as water. (4) A hydride ion is transferred from a third cysteine residue with the concomitant formation of a disulfide bond. (5) The C-3′ radical recaptures the originally abstracted hydrogen atom. (6) An electron is transferred from R2 to reduce the thiyl radical, which also accepts a proton. The deoxyribonucleotide is free to leave R1. The disulfide formed in the active site must be reduced to begin another cycle.

1. The reaction begins with the transfer of an electron from a cysteine residue on R1 to the tyrosyl radical on R2. The loss of an electron generates a highly reactive *cysteine thiyl radical* within the active site of R1.

2. This radical then abstracts a hydrogen atom from C-3′ of the ribose unit, generating a radical at that carbon atom.

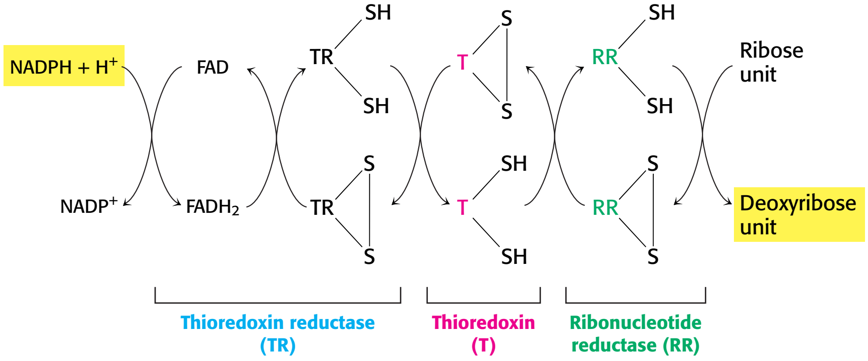
3. The radical at C-3′ promotes the release of the OH^- from the C-2′ carbon atom. Protonated by a second cysteine residue, the departing OH^- leaves as a water molecule.

4. A hydride ion (a proton with two electrons) is then transferred from a third cysteine residue to complete the reduction of the position, form a disulfide bond, and re-form a radical.

5. This C-3′ radical recaptures the same hydrogen atom originally abstracted by the first cysteine residue, and the deoxyribonucleotide is free to leave the enzyme.

6. R2 provides an electron to reduce the thiyl radical. The disulfide bond generated in the enzyme's active site must then be reduced to regenerate the active enzyme.

The electrons for this reduction come from NADPH, but not directly. One carrier of reducing power linking NADPH with the reductase is *thioredoxin,* a 12-kDa protein with two exposed cysteine residues near each other. These sulfhydryls are oxidized to a disulfide in the reaction catalyzed by ribonucleotide reductase itself. In turn, reduced thioredoxin is regenerated by electron flow from NADPH. This reaction is catalyzed by *thioredoxin reductase,* a flavoprotein. Electrons flow from NADPH to bound FAD of the reductase, to the disulfide of oxidized thioredoxin, and then to ribonucleotide reductase and finally to the ribose unit.



Stable radicals other than tyrosyl radical are employed by other ribonucleotide reductases

Ribonucleotide reductases that do not contain tyrosyl radicals have been characterized in other organisms. Instead, these enzymes contain other stable radicals that are generated by other processes. For example, in one class of reductases, the coenzyme adenosylcobalamin (vitamin B_{12}) is the radical source (Section 22.3). Despite differences in the stable radical employed, the active sites of these enzymes are similar to that of the *E. coli* ribonucleotide reductase, and they appear to act by the same mechanism, based on the exceptional reactivity of cysteine radicals. Thus, these enzymes have a common ancestor but evolved a range of mechanisms for generating stable radical species that function well under different growth conditions. The primordial enzymes appear to have been inactivated by oxygen, whereas enzymes such as the *E. coli* enzyme make use of oxygen to generate the initial tyrosyl radical. Note that the reduction of ribonucleotides to deoxyribonucleotides is a difficult reaction chemically, likely to require a sophisticated catalyst. The existence of a common protein enzyme framework for this process strongly suggests that proteins joined the RNA world before the evolution of DNA as a stable storage form for genetic information.

Thymidylate is formed by the methylation of deoxyuridylate

Uracil, produced by the pyrimidine synthesis pathway, is not a component of DNA. Rather, DNA contains *thymine,* a methylated analog of uracil. Another step is required to generate thymidylate from uracil. *Thymidylate synthase* catalyzes this finishing touch: deoxyuridylate (dUMP) is methylated to thymidylate (TMP). Recall that thymidylate synthase also functions in the thymine salvage pathway (p. 748). As will be described in Chapter 28, the methylation of this nucleotide marks sites of DNA damage for repair and, hence, helps preserve the integrity of the genetic information stored in DNA. The methyl donor in this reaction is N^5, N^{10}-methylenetetrahydrofolate rather than S-adenosylmethionine (Section 24.2).

The methyl group becomes attached to the C-5 atom within the aromatic ring of dUMP, but this carbon atom is not a good nucleophile and

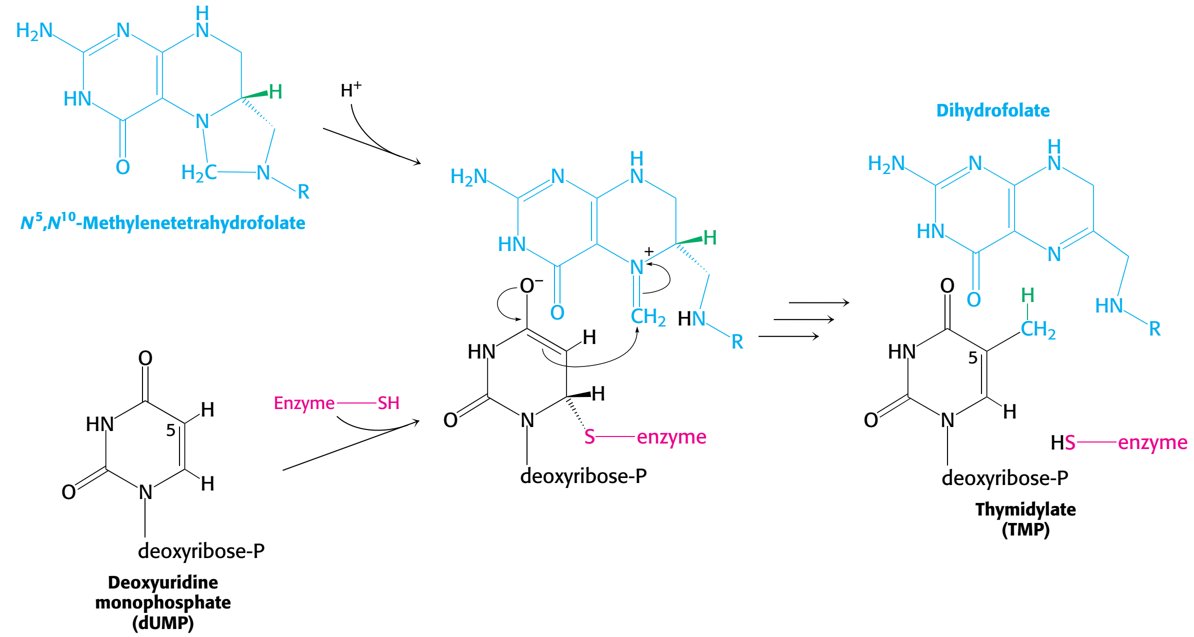
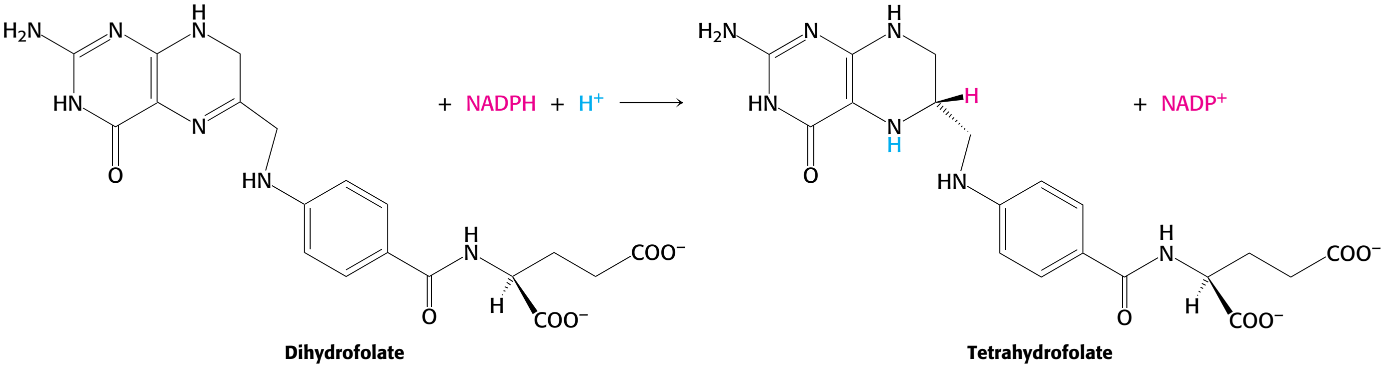

H2N
N
H
N
HN
H
N
O
H2C
N
R
N5,N10-Methylenetetrahydrofolate
H+
Dihydrofolate
O
HN
5
H
O
N
H
deoxyribose-P
Deoxyuridine
monophosphate
(dUMP)
Enzyme—SH
O−
CH2
HN
R
S—enzyme
deoxyribose-P
CH2
HS—enzyme
Thymidylate
(TMP)

FIGURE 25.12 Thymidylate synthesis. Thymidylate synthase catalyzes the addition of a methyl group (derived from N^5, N^{10}-methylenetetrahydrofolate) to dUMP to form TMP. The addition of a thiolate from the enzyme activates dUMP. Opening the five-membered ring of the THF derivative prepares the methylene group for nucleophilic attack by the activated dUMP. The reaction is completed by the transfer of a hydride ion to form dihydrofolate.

cannot itself attack the appropriate group on the methyl donor. Thymidylate synthase promotes methylation by adding a thiolate from a cysteine side chain to this ring to generate a nucleophilic species that can attack the methylene group of N^5, N^{10}-methylenetetrahydrofolate (Figure 25.12). This methylene group, in turn, is activated by distortions imposed by the enzyme that favor opening the five-membered ring. The activated dUMP's attack on the methylene group forms the new carbon–carbon bond. The intermediate formed is then converted into product: a hydride ion is transferred from the tetrahydrofolate ring to transform the methylene group into a methyl group, and a proton is abstracted from the carbon atom bearing the methyl group to eliminate the cysteine and regenerate the aromatic ring. The tetrahydrofolate derivative loses both its methylene group and a hydride ion and, hence, is oxidized to dihydrofolate. For the synthesis of more thymidylate, tetrahydrofolate must be regenerated.

Dihydrofolate reductase catalyzes the regeneration of tetrahydrofolate, a one-carbon carrier

Tetrahydrofolate is regenerated from the dihydrofolate that is produced in the synthesis of thymidylate. This regeneration is accomplished by *dihydrofolate reductase* with the use of NADPH as the reductant.


+ NADPH + H+
+ NADP+
COO−
H
Dihydrofolate
Tetrahydrofolate

A hydride ion is directly transferred from the nicotinamide ring of NADPH to the pteridine ring of dihydrofolate. The bound dihydrofolate and NADPH are held in close proximity to facilitate the hydride transfer.

Several valuable anticancer drugs block the synthesis of thymidylate

Rapidly dividing cells require an abundant supply of thymidylate for the synthesis of DNA. The vulnerability of these cells to the inhibition of TMP synthesis has been exploited in the treatment of cancer. Thymidylate synthase and dihydrofolate reductase are choice targets of chemotherapy (Figure 25.13).

Fluorouracil, an anticancer drug, is converted in vivo into *fluorodeoxyuridylate* (F-dUMP). This analog of dUMP irreversibly inhibits thymidylate synthase after acting as a normal substrate through part of the catalytic cycle. Recall that the formation of TMP requires the removal of a proton (H^+) from C-5 of the bound nucleotide (Figure 25.12). However, the enzyme cannot abstract F^+ from F-dUMP, and so catalysis is blocked at the stage of the covalent complex formed by F-dUMP, methylenetetrahydrofolate, and the sulfhydryl group of the enzyme (Figure 25.14). We see here an example of *suicide inhibition*, in which an enzyme converts a substrate into a reactive inhibitor that halts the enzyme's catalytic activity (Section 8.5).

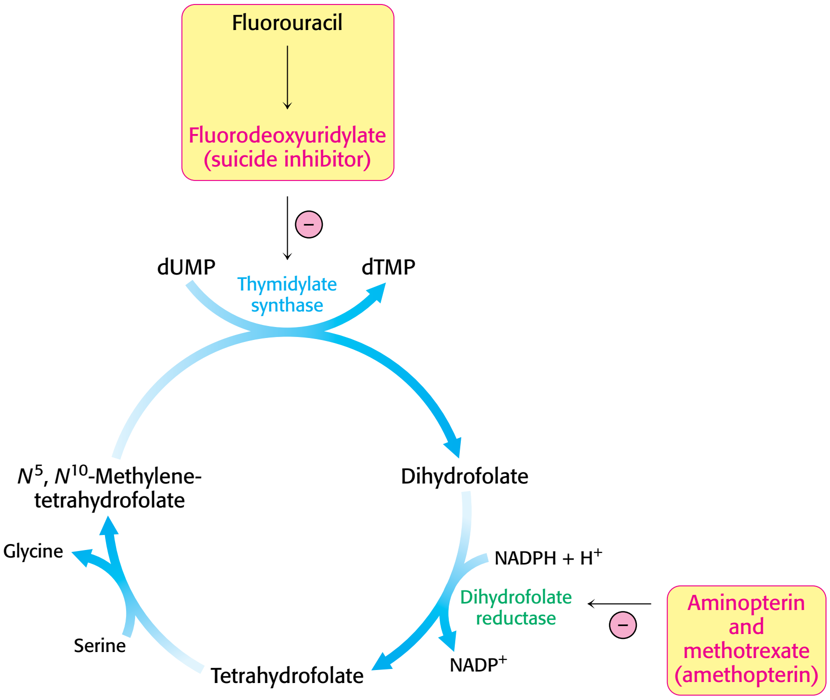
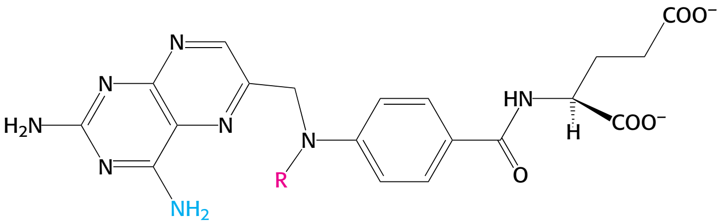

Fluorouracil
Fluorodeoxyuridylate (suicide inhibitor)
–
dUMP
Thymidylate synthase
dTMP
Dihydrofolate
NADPH + H+
Dihydrofolate reductase
NADP+
–
Aminopterin and methotrexate (amethopterin)
Tetrahydrofolate
Serine
Glycine
N5, N10-Methylene-tetrahydrofolate

FIGURE 25.13 Anticancer drug targets. Thymidylate synthase and dihydrofolate reductase are choice targets in cancer chemotherapy because the generation of large quantities of precursors for DNA synthesis is required for rapidly dividing cancer cells.

The synthesis of TMP can also be blocked by inhibiting the regeneration of tetrahydrofolate. Analogs of dihydrofolate, such as *aminopterin* and *methotrexate* (amethopterin), are potent competitive inhibitors ($K_i < 1$ nM) of dihydrofolate reductase.


COO–
N
N
H2N
N
N
N
R
HN
H
COO–
O
NH2

Aminopterin (R = H) or methotrexate (R = CH_3)

Methotrexate is a valuable drug in the treatment of many rapidly growing tumors, such as those in acute leukemia and choriocarcinoma, a cancer derived from placental cells. However, methotrexate kills rapidly replicating

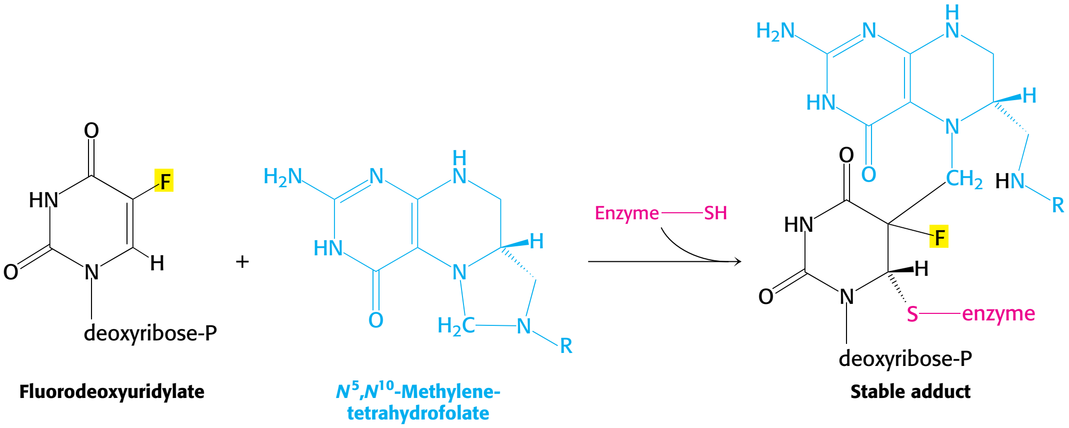
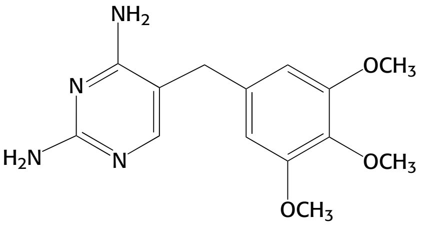

O
F
HN
O
N
H
deoxyribose-P
+
H2N
N
H
N
HN
N
H
O
H2C
N
R
Enzyme—SH
H2N
N
H
N
HN
N
H
O
O
CH2
HN
R
HN
F
H
O
N
S—enzyme
deoxyribose-P
Fluorodeoxyuridylate
N5,N10-Methylene-tetrahydrofolate
Stable adduct

FIGURE 25.14 Suicide inhibition. Fluorodeoxyuridylate (generated from fluorouracil) traps thymidylate synthase in a form that cannot proceed down the reaction pathway.



Trimethoprim

cells whether they are malignant or not. Stem cells in bone marrow, epithelial cells of the intestinal tract, and hair follicles are vulnerable to the action of this folate antagonist, accounting for its toxic side effects, which include weakening of the immune system, nausea, and hair loss.

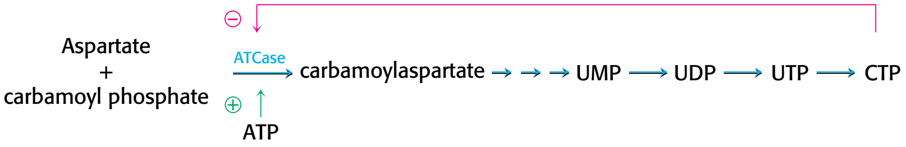
Folate analogs such as *trimethoprim* have potent antibacterial and antiprotozoal activity. Trimethoprim binds 10^5-fold less tightly to mammalian dihydrofolate reductase than it does to reductases of susceptible microorganisms. Small differences in the active-site clefts of these enzymes account for the highly selective antimicrobial action. The combination of trimethoprim and sulfamethoxazole (an inhibitor of folate synthesis) is widely used to treat infections such as bronchitis, traveler's diarrhea, and urinary tract infections.

25.4 Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition

Nucleotide biosynthesis is regulated by feedback inhibition in a manner similar to the regulation of amino acid biosynthesis (Section 24.3). These regulatory pathways ensure that the various nucleotides are produced in the required quantities.

Pyrimidine biosynthesis is regulated by aspartate transcarbamoylase

Aspartate transcarbamoylase, one of the key enzymes for the regulation of pyrimidine biosynthesis in bacteria, was described in detail in Chapter 10. Recall that *ATCase is inhibited by CTP, the final product of pyrimidine biosynthesis,* and stimulated by ATP.


Aspartate
+
carbamoyl phosphate
ATCase
⊖
⊕
ATP
carbamoylaspartate
UMP
UDP
UTP
CTP

Carbamoyl phosphate synthetase II is also a site of feedback inhibition in both prokaryotes and eukaryotes.

The synthesis of purine nucleotides is controlled by feedback inhibition at several sites

The regulatory scheme for purine nucleotides is more complex than that for pyrimidine nucleotides (Figure 25.15).

1. The committed step in purine nucleotide biosynthesis is the conversion of PRPP into phosphoribosylamine by *glutamine phosphoribosyl amidotransferase.* This important enzyme is feedback-inhibited by many purine ribonucleotides. It is noteworthy that AMP and GMP, the final products of the pathway, synergistically inhibit the amidotransferase.

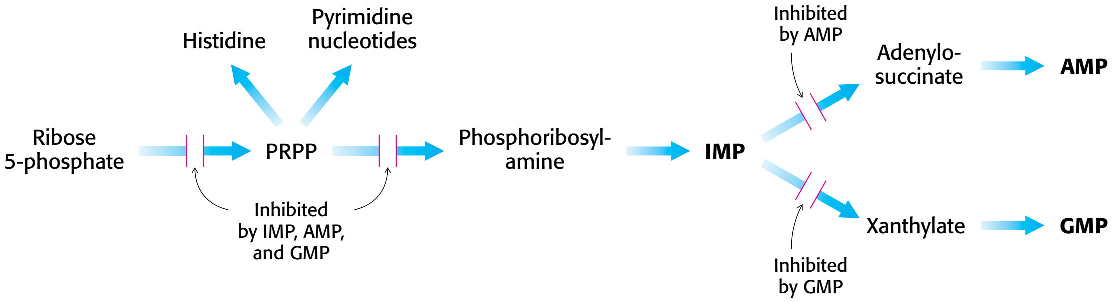

Histidine
Pyrimidine nucleotides
Ribose 5-phosphate
PRPP
Inhibited by IMP, AMP, and GMP
Phosphoribosyl-amine
IMP
Inhibited by AMP
Adenylo-succinate
AMP
Xanthylate
GMP
Inhibited by GMP

FIGURE 25.15 Control of purine biosynthesis. Feedback inhibition controls both the overall rate of purine biosynthesis and the balance between AMP and GMP production.

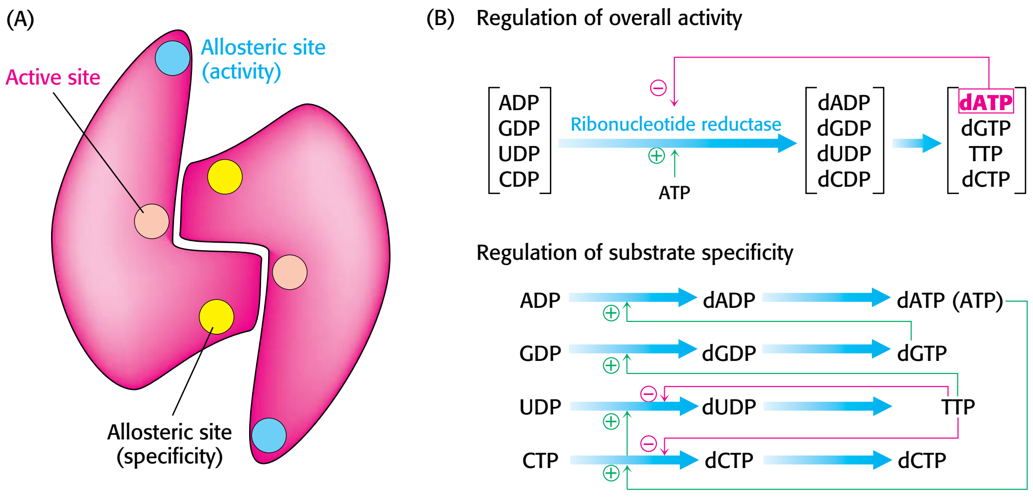

(A)
Active site
Allosteric site (activity)
Allosteric site (specificity)
(B) Regulation of overall activity
ADP
GDP
UDP
CDP
Ribonucleotide reductase
⊖
⊕
ATP
dADP
dGDP
dUDP
dCDP
dATP
dGTP
TTP
dCTP
Regulation of substrate specificity
ADP
dADP
dATP (ATP)
GDP
dGDP
dGTP
UDP
dUDP
TTP
CTP
dCTP
dCTP

FIGURE 25.16 Regulation of ribonucleotide reductase. (A) Each subunit in the R1 dimer contains two allosteric sites in addition to the active site. One site regulates the overall activity and the other site regulates substrate specificity. (B) The patterns of regulation with regard to different nucleoside diphosphates demonstrated by ribonucleotide reductase.

2. Inosinate is the branch point in the synthesis of AMP and GMP. *The reactions leading away from inosinate are sites of feedback inhibition.* AMP inhibits the conversion of inosinate into adenylosuccinate, its immediate precursor. Similarly, GMP inhibits the conversion of inosinate into xanthylate, its immediate precursor.

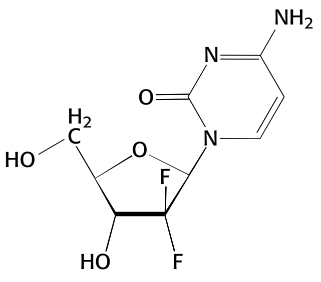
3. As already noted, GTP is a substrate in the synthesis of AMP, whereas ATP is a substrate in the synthesis of GMP (Figure 25.7). This *reciprocal substrate relation* tends to balance the synthesis of adenine and guanine ribonucleotides.

Note that the synthesis of PRPP by PRPP synthetase is highly regulated even though it is not the committed step in purine synthesis. Mutations have been identified in PRPP synthetase that result in a loss of allosteric response to nucleotides without any effect on catalytic activity of the enzyme. A consequence of this mutation is an overabundance of purine nucleotides that can result in gout, a pathological condition discussed later in the chapter.

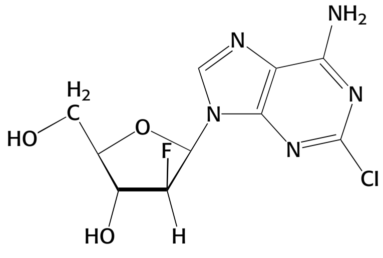
The synthesis of deoxyribonucleotides is controlled by the regulation of ribonucleotide reductase

The reduction of ribonucleotides to deoxyribonucleotides is precisely controlled by allosteric interactions. Each polypeptide of the R1 subunit of the aerobic *E. coli* ribonucleotide reductase contains two allosteric sites: one of them controls the *overall activity* of the enzyme, and the other regulates *substrate specificity* (Figure 25.16A). The overall catalytic activity of ribonucleotide reductase is diminished by the binding of dATP, which signals an abundance of deoxyribonucleotides (Figure 25.16B). The binding of ATP reverses this feedback inhibition. The binding of dATP or ATP to the substrate-specificity control site enhances the reduction of UDP and CDP, the pyrimidine nucleotides. The binding of thymidine triphosphate (TTP) promotes the reduction of GDP and inhibits the further reduction of pyrimidine ribonucleotides. The subsequent increase in the level of dGTP stimulates the reduction of ATP to dATP. This complex pattern of regulation supplies the appropriate balance of the four deoxyribonucleotides needed for the synthesis of DNA.

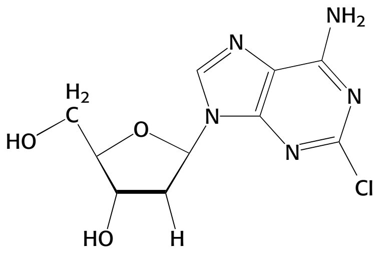
Ribonucleotide reductase is an attractive target for cancer therapy and a number of clinically approved anticancer drugs mimic substrates and regulators of the enzyme. The pyrimidine analog gemcitabine, once converted to the diphosphate form in vivo, becomes a suicide inhibitor


NH2
N
O
N
H2
C
HO
O
F
HO
F

Gemcitabine


NH2
N
N
N
N
Cl
H2
C
HO
O
F
HO
H

Clofarabine


NH2
N
N
N
N
Cl
H2
C
HO
O
HO
H

Cladribine

of the reductase used to treat advanced pancreatic cancer. The purine analogs clofarabine and cladribine, upon conversion to their triphosphate forms in vivo, are dATP analogs and act as allosteric inhibitors of the enzyme. Clofarabine is used to treat pediatric acute myeloid leukemia while cladribine is effective against some forms of chronic lymphoid leukemia.

25.5 Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions

Nucleotides are vital to a host of biochemical processes. It is not surprising, then, that disruption of nucleotide metabolism would have a variety of physiological effects. The nucleotides of a cell undergo continual turnover. Nucleotides are hydrolytically degraded to nucleosides by *nucleotidases.* The phosphorolytic cleavage of nucleosides to free bases and ribose 1-phosphate (or deoxyribose 1-phosphate) is catalyzed by *nucleoside phosphorylases.* Ribose 1-phosphate is isomerized by *phosphoribomutase* to ribose 5-phosphate, a substrate in the synthesis of PRPP. Some of the bases are reused to form nucleotides by salvage pathways. Others are degraded to products that are excreted (Figure 25.17). A deficiency of an enzyme can disrupt these pathways, leading to a pathological condition.

The loss of adenosine deaminase activity results in severe combined immunodeficiency

The pathway for the degradation of AMP includes an extra step because adenosine is not a substrate for nucleoside phosphorylase. First, the phosphate is removed by a nucleotidase to yield the nucleoside adenosine (Figure 25.17). In the extra step, adenosine is deaminated by *adenosine deaminase* to form inosine.

A deficiency in adenosine deaminase activity is associated with some forms of *severe combined immunodeficiency* (SCID), an immunological disorder.

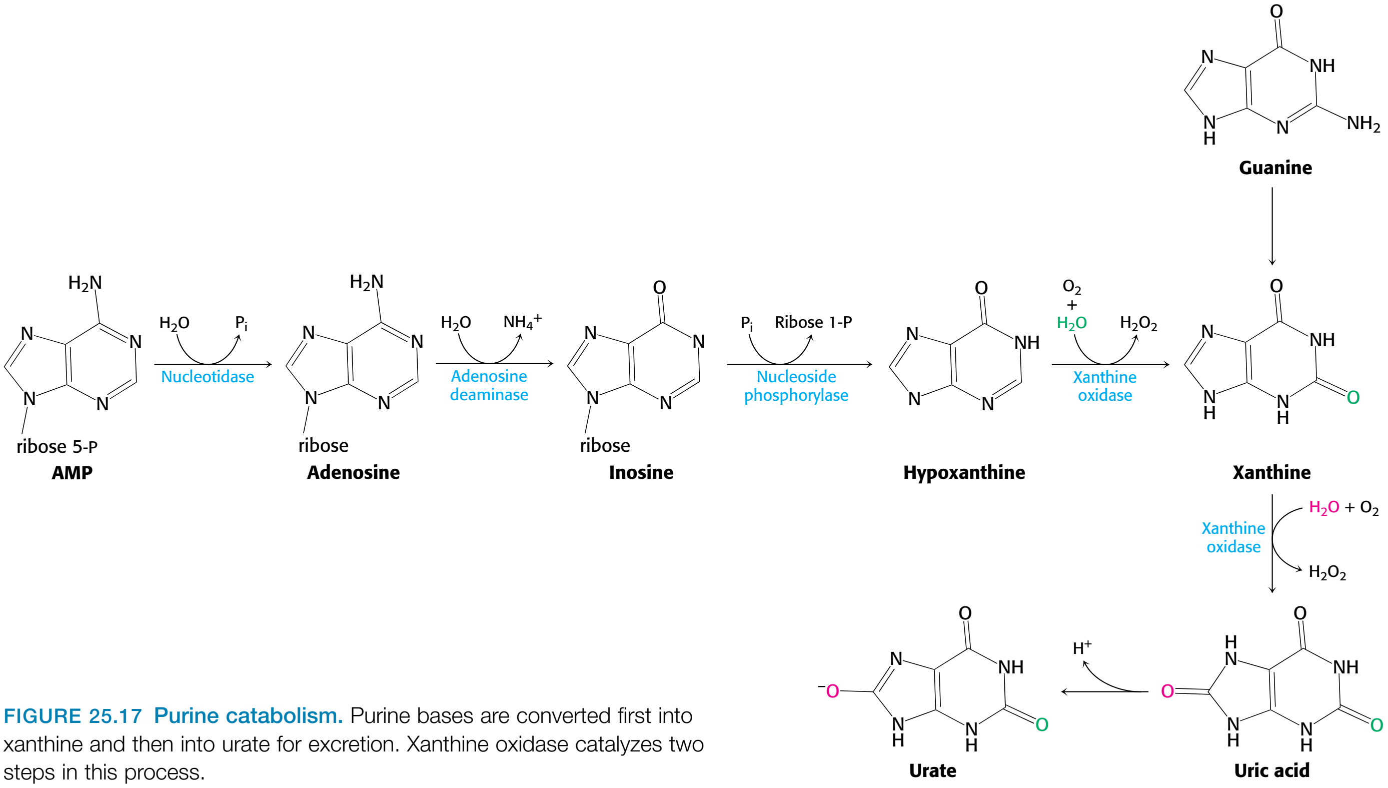

Guanine
AMP
ribose 5-P
H_2O
P_i
Nucleotidase
Adenosine
ribose
H_2O
NH_4^+
Adenosine deaminase
Inosine
ribose
P_i
Ribose 1-P
Nucleoside phosphorylase
Hypoxanthine
O_2 + H_2O
H_2O_2
Xanthine oxidase
Xanthine
Xanthine oxidase
$H_2O + O_2$
H_2O_2
Uric acid
H^+
Urate

FIGURE 25.17 **Purine catabolism.** Purine bases are converted first into xanthine and then into urate for excretion. Xanthine oxidase catalyzes two steps in this process.

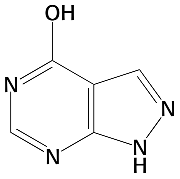
Persons with the disorder have acute recurring infections, often leading to death at an early age. SCID is characterized by a loss of T cells, which are crucial to the immune response (Section 34.5). Although the biochemical basis of the disorder is not clearly established, a lack of adenosine deaminase results in an increase of 50 to 100 times the normal level of dATP, which inhibits ribonucleotide reductase and, consequently, DNA synthesis. Moreover, adenosine itself is a powerful signal molecule with a role in a number of regulatory pathways. Disruption in the levels of adenosine may also be deleterious. SCID is often called the "bubble boy disease" because its treatment may include complete isolation of the patient from the environment. Adenosine deaminase deficiency has been successfully treated by gene therapy.

Gout is induced by high serum levels of urate

Inosine generated by adenosine deaminase is subsequently metabolized by *nucleoside phosphorylase* to hypoxanthine. *Xanthine oxidase,* a molybdenum- and iron-containing flavoprotein, oxidizes hypoxanthine to *xanthine* and then to *uric acid.* Molecular oxygen, the oxidant in both reactions, is reduced to H_2O_2, which is decomposed to H_2O and O_2 by catalase. Uric acid loses a proton at physiological pH to form *urate. In human beings, urate is the final product of purine degradation and is excreted in the urine.*

High serum levels of urate (hyperuricemia) induce the painful joint disease *gout*. In this disease, the sodium salt of urate crystallizes in the fluid and lining of the joints. The small joint at the base of the big toe is a common site for sodium urate buildup, although the salt accumulates at other joints also. Painful inflammation results when cells of the immune system engulf the sodium urate crystals. The kidneys, too, may be damaged by the deposition of urate crystals. Gout is a common medical problem, affecting 1% of the population of Western countries. It is nine times as common in men as in women.

Administration of *allopurinol,* an analog of hypoxanthine, is one treatment for gout. The mechanism of action of allopurinol is interesting: it acts *first as a substrate and then as an inhibitor* of xanthine oxidase. The oxidase hydroxylates allopurinol to *alloxanthine (oxipurinol),* which then remains tightly bound to the active site. The binding of alloxanthine keeps the molybdenum atom of xanthine oxidase in the +4 oxidation state instead of it returning to the +6 oxidation state as in a normal catalytic cycle. We see here another example of *suicide inhibition.*



Allopurinol

The synthesis of urate from hypoxanthine and xanthine decreases soon after the administration of allopurinol. The serum concentrations of hypoxanthine and xanthine rise, and that of urate drops.

The average serum level of urate in human beings is close to the solubility limit and is higher than levels found in other primates. What is the selective advantage of a urate level so high that it teeters on the brink of gout in many people? It turns out that urate has a markedly beneficial action. Urate is a highly effective scavenger of reactive oxygen species. Indeed, urate is about as effective as ascorbate (vitamin C) as an antioxidant. The increased level of urate in human beings may protect against reactive oxygen species that are implicated in a host of pathological conditions (Table 18.3).

Lesch–Nyhan syndrome is a dramatic consequence of mutations in a salvage-pathway enzyme

Mutations in genes that encode nucleotide biosynthetic enzymes can reduce levels of needed nucleotides and can lead to an accumulation of intermediates. A nearly total absence of hypoxanthine-guanine phosphoribosyltransferase (HGPRT) has unexpected and devastating consequences. The most striking expression of this inborn error of metabolism, called the

Lesch–Nyhan syndrome, is *compulsive self-destructive behavior.* At age 2 or 3, children with this disease begin to bite their fingers and lips and will chew them off if unrestrained. These children also behave aggressively toward others. *Mental deficiency* and *spasticity* are other characteristics of the Lesch–Nyhan syndrome. Elevated levels of urate in the serum lead to the formation of kidney stones early in life, followed by the symptoms of gout years later. The disease is inherited as a sex-linked recessive disorder.

What is the connection between the absence of HGPRT activity and the behavioral characteristics of Lesch-Nyhan syndrome? The answer is not clear, but it is possible to speculate. The brain has limited capacity for de novo purine synthesis. Consequently, lack of HGPRT results in a deficiency of purine nucleotides. ATP and ADP, formed from inosinate, are especially important in the brain as signal molecules. These nucleotides bind to and activate G-protein coupled receptors that regulate the dopamine secreting neurons. Thus, the lack of HGPRT results in an imbalance of key neurotransmitters. Moreover, the guanosine nucleotides required for G-protein function may also be in short supply. The Lesch–Nyhan syndrome demonstrates that the salvage pathway for the synthesis of IMP and GMP is not of minor importance. Moreover, the Lesch–Nyhan syndrome reveals that *abnormal behavior such as self-mutilation and extreme hostility can be caused by the absence of a single enzyme.* Psychiatry will no doubt benefit from the unraveling of the molecular basis of such mental disorders.

Folic acid deficiency promotes birth defects such as spina bifida

Spina bifida is one of a class of birth defects characterized by the incomplete or incorrect formation of the neural tube early in development. In the United States, the prevalence of *neural-tube defects* is approximately 1 case per 1000 births. A variety of studies have demonstrated that the prevalence of neural-tube defects is reduced by as much as 70% when women take folic acid as a dietary supplement before and during the first trimester of pregnancy. One hypothesis is that more folate derivatives are needed for the synthesis of DNA precursors when cell division is frequent and substantial amounts of DNA must be synthesized.

SUMMARY

25.1 The Pyrimidine Ring Is Assembled de Novo or Recovered by Salvage Pathways

The pyrimidine ring is assembled first and then linked to ribose phosphate to form a pyrimidine nucleotide. 5-Phosphoribosyl-1-pyrophosphate is the donor of the ribose phosphate moiety. The synthesis of the pyrimidine ring starts with the formation of carbamoylaspartate from carbamoyl phosphate and aspartate, a reaction catalyzed by aspartate transcarbamoylase. Dehydration, cyclization, and oxidation yield orotate, which reacts with PRPP to give orotidylate. Decarboxylation of this pyrimidine nucleotide yields UMP. CTP is then formed by the amination of UTP.

25.2 Purine Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways

The purine ring is assembled from a variety of precursors: glutamine, glycine, aspartate, N^{10}-formyltetrahydrofolate, and CO_2. The committed step in the de novo synthesis of purine nucleotides is the formation of 5-phosphoribosylamine from PRPP and glutamine. The purine ring is assembled on ribose phosphate, in contrast with the de novo synthesis of pyrimidine nucleotides. The addition of glycine, followed by formylation, amination, and ring closure, yields 5-aminoimidazole ribonucleotide. This intermediate contains the completed five-membered ring of the purine skeleton. The addition of CO_2, the nitrogen atom of aspartate, and a formyl group, followed by ring closure, yields inosinate, a purine ribonucleotide. AMP and GMP are formed from IMP. Purine ribonucleotides can also be synthesized by a salvage pathway in which a preformed base reacts directly with PRPP.

25.3 Deoxyribonucleotides Are Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism

Deoxyribonucleotides, the precursors of DNA, are formed in *E. coli* by the reduction of ribonucleoside diphosphates. These conversions are catalyzed by ribonucleotide reductase. Electrons are transferred from NADPH to sulfhydryl groups at the active sites of this enzyme by thioredoxin. A tyrosyl free radical generated by an iron center in the reductase initiates a radical reaction on the sugar, leading to the exchange of H for OH at C-2′. TMP is formed by the methylation of dUMP. The donor of a methylene group and a hydride in this reaction is N^5, N^{10}-methylenetetrahydrofolate, which is converted into dihydrofolate. Tetrahydrofolate is regenerated by the reduction of dihydrofolate by NADPH. Dihydrofolate reductase, which catalyzes this reaction, is inhibited by folate analogs such as aminopterin and methotrexate. These compounds and fluorouracil, an inhibitor of thymidylate synthase, are used as anticancer drugs.

25.4 Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition

Pyrimidine biosynthesis in *E. coli* is regulated by the feedback inhibition of aspartate transcarbamoylase, the enzyme that catalyzes the committed step. CTP inhibits and ATP stimulates this enzyme. The feedback inhibition of glutamine-PRPP amidotransferase by purine nucleotides is important in regulating their biosynthesis. Disruption of the regulation and activity of ribonucleotide reductase by drugs is an effective chemotherapy for some types of cancer.

25.5 Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions

Severe combined immunodeficiency results from the absence of adenosine deaminase, an enzyme in the purine degradation pathway. Purines are degraded to urate in human beings. Gout, a disease that affects joints and leads to arthritis, is associated with an excessive accumulation of urate. The Lesch–Nyhan syndrome, a genetic disease characterized by self-mutilation, mental deficiency, and gout, is caused by the absence of hypoxanthine-guanine phosphoribosyltransferase. This enzyme is essential for the synthesis of purine nucleotides by the salvage pathway. Neural-tube defects are more frequent when a pregnant woman is deficient in folate derivatives early in pregnancy, possibly because of the important role of these derivatives in the synthesis of DNA precursors.

KEY TERMS

pyrimidine nucleotide (p. 744)
carbamoyl phosphate synthetase II (CPS II) (p. 745)
ATP-grasp fold (p. 745)
5-phosphoribosyl-1-pyrophosphate (PRPP) (p. 746)
orotidylate (p. 747)
salvage pathway (p. 748)
purine nucleotide (p. 748)
glutamine phosphoribosyl amidotransferase (p. 749)
ribonucleotide reductase (p. 753)
thymidylate synthase (p. 755)
dihydrofolate reductase (p. 756)
severe combined immunodeficiency (SCID) (p. 760)
gout (p. 761)
Lesch–Nyhan syndrome (p. 761)
spina bifida (p. 762)
neural-tube defect (p. 762)

PROBLEMS

1. *From the beginning or extract and save and reuse.* Differentiate between the de novo synthesis of nucleotides and salvage-pathway synthesis.

2. *Finding their roots 1.* Identify the source of the atoms in the pyrimidine ring

3. *Finding their roots 2.* Identify the source of the atoms in the purine ring.

4. *Multifaceted.* List some of the biochemical roles played by nucleotides.

5. *An s instead of a t?* Differentiate between a nucleoside and a nucleotide.

6. *Associate 'em.*

(a) Excessive urate ______
(b) Lack of adenosine deaminase ______
(c) Lack of HGPRT ______
(d) Carbamoyl phosphate ______
(e) Inosinate ______
(f) Ribonucleotide reductase ______
(g) Lack of folic acid ______
(h) Glutamine phosphoribosyl transferase ______
(i) Single ring ______
(j) Bicyclic ring ______
(k) Precursor to CTP ______

1. Spina bifida
2. Precursor to both ATP and GTP
3. Purine
4. Deoxynucleotide synthesis
5. UTP
6. Lesch-Nyhan disease
7. Immunodeficiency
8. Pyrimidine
9. Gout
10. First step in pyrimidine synthesis
11. Committed step in purine synthesis

7. *Safe passage.* What is substrate channeling? How does it affect enzyme efficiency?

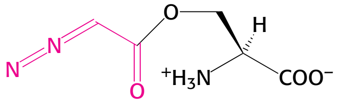
8. *Activated ribose phosphate.* Write a balanced equation for the synthesis of PRPP from glucose through the oxidative branch of the pentose phosphate pathway.

9. *Making a pyrimidine.* Write a balanced equation for the synthesis of orotate from glutamine, CO_2, and aspartate.

10. *Identifying the donor.* What is the activated reactant in the biosynthesis of each of the following compounds?

(a) Phosphoribosylamine
(b) Carbamoylaspartate
(c) Orotidylate (from orotate)
(d) Phosphoribosylanthranilate

11. *Inhibiting purine biosynthesis.* Amidotransferases are inhibited by the antibiotic azaserine (*O*-diazoacetyl-L-serine), which is an analog of glutamine.

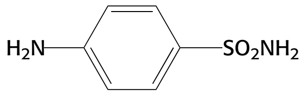


Azaserine

Which intermediates in purine biosynthesis would accumulate in cells treated with azaserine?

12. *The price of methylation.* Write a balanced equation for the synthesis of TMP from dUMP that is coupled to the conversion of serine into glycine.

13. *Sulfa action.* Bacterial growth is inhibited by sulfanilamide and related sulfa drugs, and there is a concomitant accumulation of 5-aminoimidazole-4-carboxamide ribonucleotide. This inhibition is reversed by the addition of *p*-aminobenzoate.



Sulfanilamide

Propose a mechanism for the inhibitory effect of sulfanilamide.

14. *HAT medium.* Mutant cells unable to synthesize nucleotides by salvage pathways are very useful tools in molecular and cell biology. Suppose that cell A lacks thymidine kinase, the enzyme catalyzing the phosphorylation of thymidine to thymidylate, and that cell B lacks hypoxanthine-guanine phosphoribosyl transferase.

(a) Cell A and cell B do not proliferate in a HAT medium containing *h*ypoxanthine, *a*minopterin or *a*methopterin (methotrexate), and *t*hymine. However, cell C, formed by the fusion of cells A and B, grows in this medium. Why?

(b) Suppose that you want to introduce foreign genes into cell A. Devise a simple means of distinguishing between cells that have taken up foreign DNA and those that have not.

15. *Vitamin requirement.* What is the role of folate in pyrimidine synthesis and what is the consequence of a folate deficiency during development?

16. *Bringing equilibrium.* What is the reciprocal substrate relation in the synthesis of ATP and GTP?

17. *Find the label.* Suppose that cells are grown on amino acids that have all been labeled at the α carbons with ^{13}C. Identify the atoms in cytosine and guanine that will be labeled with ^{13}C.

18. *Need a map?* Describe the path for the synthesis of TTP from UTP.

19. *Adjunct therapy.* Allopurinol is sometimes given to patients with acute leukemia who are being treated with anticancer drugs. Why is allopurinol used?

20. *A hobbled enzyme.* Both side-chain oxygen atoms of aspartate 27 at the active site of dihydrofolate reductase form hydrogen bonds with the pteridine ring of folates. The

importance of this interaction was assessed by studying two mutants at this position, Asn 27 and Ser 27. The dissociation constant of methotrexate was 0.07 nM for the wild type, 1.9 nM for the Asn 27 mutant, and 210 nM for the Ser 27 mutant, at 25°C. Calculate the standard free energy of the binding of methotrexate by these three proteins. What is the decrease in binding energy resulting from each mutation?

21. *Needed supplies*. Why are cancer cells especially sensitive to inhibitors of TMP synthesis?

22. *Alternative routes*. Orotic aciduria results when one of the enzyme activities of UMP synthetase is absent. This syndrome is characterized by large amounts of orotic acid in the blood and urine, megaloblastic anemia (characterized by large, immature, and dysfunctional red blood cells) and retarded growth. Suggest a possible treatment for this condition.

23. *Correcting deficiencies.* Suppose that a person is found who is deficient in an enzyme required for IMP synthesis. How might this person be treated?

24. *Labeled nitrogen.* Purine biosynthesis is allowed to take place in the presence of [^{15}N]aspartate, and the newly synthesized GTP and ATP are isolated. What positions are labeled in the two nucleotides?

25. *On the trail of carbons.* Tissue culture cells were incubated with glutamine labeled with ^{15}N in the amide group. Subsequently, IMP was isolated and found to contain some ^{15}N. Which atoms in IMP were labeled?

26. *Mechanism of action*. What is the biochemical basis of allopurinol treatment for gout?

27. *Changed inhibitor.* Xanthine oxidase treated with allopurinol results in the formation of a new compound that is an extremely potent inhibitor of the enzyme. Propose a structure for this compound.

28. *Calculate the ATP footprint*. How many molecules of ATP are required to synthesize one molecule of CTP from scratch?

29. *Blockages.* What intermediate in purine synthesis will accumulate if a strain of bacteria is lacking each of the following biochemicals?

(a) Aspartate
(b) Tetrahydrofolate
(c) Glycine
(d) Glutamine

Mechanism Problems

30. *The same and not the same.* Write out mechanisms for the conversion of phosphoribosylamine into glycinamide ribonucleotide and of xanthylate into guanylate.

31. *Closing the ring.* Propose a mechanism for the conversion of 5-formamidoimidazole-4-carboxamide ribonucleotide into inosinate.

Chapter Integration Problems

32. *Different strokes.* Human beings contain two different carbamoyl phosphate synthetase enzymes. One uses glutamine as a substrate, whereas the other uses ammonia. What are the functions of these two enzymes?

33. *A generous donor.* What major biosynthetic reactions utilize PRPP?

34. *They're everywhere!* Nucleotides play a variety of roles in the cell. Give an example of a nucleotide that acts in each of the following roles or processes.

(a) Second messenger
(b) Phosphoryl-group transfer
(c) Activation of carbohydrates
(d) Activation of acetyl groups
(e) Transfer of electrons
(f) DNA sequencing
(g) Chemotherapy
(h) Allosteric effector

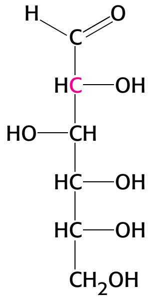
35. *Pernicious anemia.* Purine biosynthesis is impaired by vitamin B_{12} deficiency. Why? How might fatty acid and amino acid metabolism also be affected by a vitamin B_{12} deficiency?

36. *Folate deficiency*. Suppose someone was suffering from a folate deficiency. What cells would you think might be most affected? Symptoms may include diarrhea and anemia.

37. *Hyperuricemia.* Many patients with glucose 6-phosphatase deficiency have high serum levels of urate. Hyperuricemia can be induced in normal people by the ingestion of alcohol or by strenuous exercise. Propose a common mechanism that accounts for these findings.

38. *Labeled carbon.* Succinate uniformly labeled with ^{14}C is added to cells actively engaged in pyrimidine biosynthesis. Propose a mechanism by which carbon atoms from succinate could be incorporated into a pyrimidine. At what positions is the pyrimidine labeled?

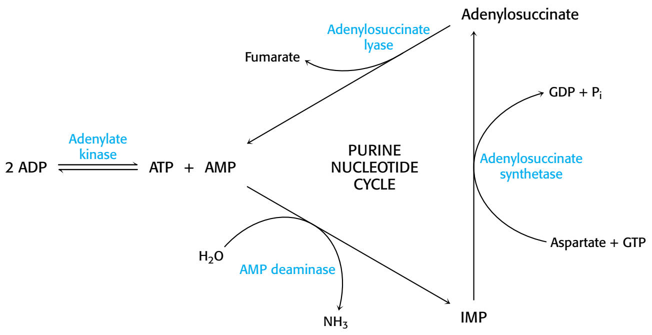
39. *Something funny going on here.* Cells were incubated with glucose labeled with ^{14}C in carbon 2, shown in red in the structure below. Later, uracil was isolated and found to contain ^{14}C in carbons 4 and 6. Account for this labeling pattern.



40. *Side effects*. Azathioprine is an immunosuppressant drug used in kidney transplants. In vivo, azathioprine is

metabolized to the hypoxanthine analog 6-mercaptopurine, which is then converted into 6-mercaptopurine ribose monophosphate, the active form of the drug. 6-Mercaptopurine ribose monophosphate also inhibits de novo purine synthesis, reducing uric acid levels in the blood and urine. However, when administered to Lesch–Nyhan patients, there was no effect on uric acid levels. Explain why this is the case.

41. *Exercising muscle.* Some interesting reactions take place in muscle tissue to facilitate the generation of ATP for contraction. In muscle contraction, ATP is converted into ADP. Adenylate kinase converts two molecules of ADP into a molecule of ATP and AMP.



(a) Why is this reaction beneficial to contracting muscle?

(b) Why is the equilibrium for the adenylate kinase approximately equal to 1?

Muscle can metabolize AMP by using the purine nucleotide cycle. The initial step in this cycle, catalyzed by AMP deaminase, is the conversion of AMP into IMP.

(c) Why might the deamination of AMP facilitate ATP formation in muscle?

(d) How does the purine nucleotide cycle assist the aerobic generation of ATP?

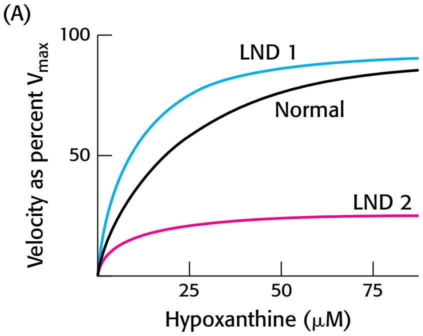
42. *A common step.* What three reactions transfer an amino group from aspartate to yield the aminated product and fumarate?

43. *Your pet duck.* You suspect that your pet duck has gout. Why should you think twice before administering a dose of allopurinol-laced bread?

Data Interpretation Problem

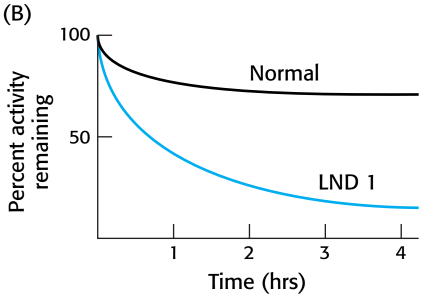
44. *Not all absences are equal.* In order to better understand the nature of the defects in HGPRT in patients suffering from Lesch–Nyhan disease, cell lines from two Lesch–Nyhan patients (LND 1 and LND 2) as well as a normal individual were established. HGPRT was then purified from each of the cell lines using the techniques examined in Chapter 3. Using the purified enzyme, the basic kinetics of the three enzymes were compared, as shown in Figure A. In each case, the same amount of purified enzyme was used in the analysis.

(a) Describe the results shown in the figure.



(b) Suggest some possible explanations for the appearance of Lesch–Nyhan disease even in the presence of apparently normal enzyme.

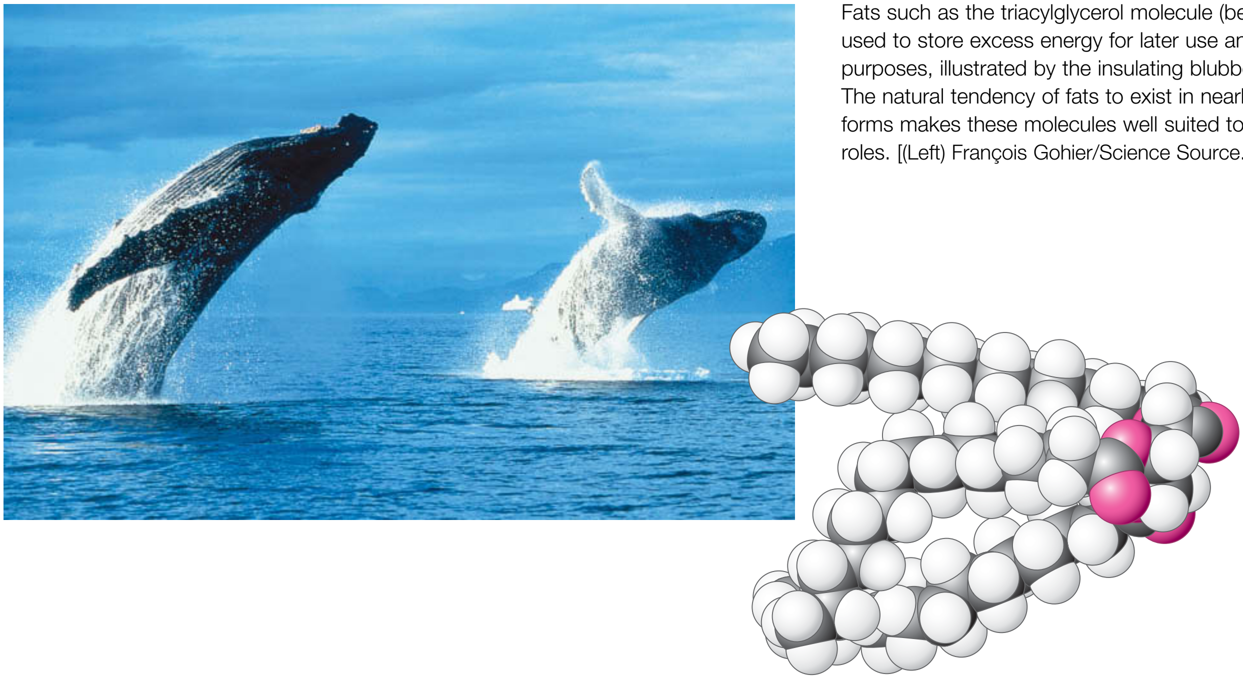
(c) The stability of the HGPRT was tested to determine if the LND 1 enzyme was more labile. The same amounts of the normal and LND 1 enzyme were incubated at 37 °C in the absence of any substrates or products. At various times, a portion of the enzyme was removed and assayed for enzyme activity. The results are shown in Figure B. Explain these results.



CHAPTER

26

The Biosynthesis of Membrane Lipids and Steroids



Fats such as the triacylglycerol molecule (below) are widely used to store excess energy for later use and to fulfill other purposes, illustrated by the insulating blubber of whales. The natural tendency of fats to exist in nearly water free forms makes these molecules well suited to these roles. [(Left) François Gohier/Science Source.]

This chapter examines the biosynthesis of three important components of biological membranes—phospholipids, sphingolipids, and cholesterol—which were introduced in Chapter 12. Triacylglycerols also are considered here because the pathway for their synthesis overlaps that of phospholipids. Cholesterol is of interest both as a membrane component and as a precursor of potent signal molecules, such as the steroid hormones progesterone, testosterone, estrogen, and cortisol.

The transport and uptake of cholesterol vividly illustrate a recurring mechanism for the entry of metabolites and signal molecules into cells. Cholesterol is transported in blood by the low-density lipoprotein (LDL) and taken up into cells by the LDL receptor on the cell surface. The LDL receptor is absent in people with *familial hypercholesterolemia*, a genetic disease. People lacking the receptor have markedly elevated cholesterol levels in the blood and cholesterol deposits on blood vessels, and they are prone to childhood heart attacks. Indeed, cholesterol is implicated in the development of atherosclerosis in people who do not have genetic defects. Thus, the regulation of cholesterol synthesis and transport can be a source of especially clear insight into the role that our understanding of biochemistry plays in medicine.

OUTLINE

26.1 Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols

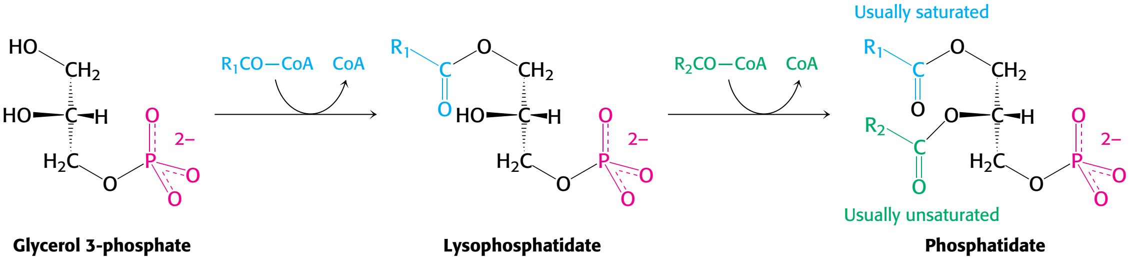
26.2 Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages

26.3 The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels

26.4 Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones

26.1 Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols

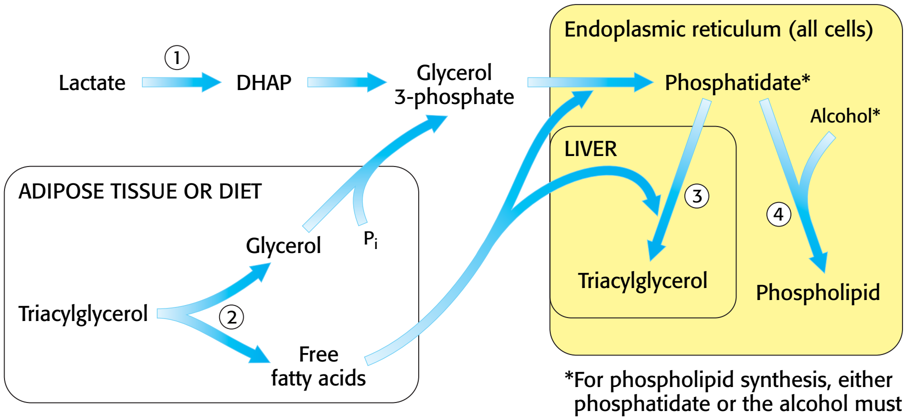
Lipid synthesis requires the coordinated action of gluconeogenesis and fatty acid metabolism, as illustrated in Figure 26.1. The common step in the synthesis of both phospholipids for membranes and triacylglycerols for energy storage is the formation of *phosphatidate* (diacylglycerol 3-phosphate). In mammalian cells, phosphatidate is synthesized in the endoplasmic reticulum and the outer mitochondrial membrane. The pathway begins with *glycerol 3-phosphate,* which is formed primarily by the reduction of dihydroxyacetone phosphate (DHAP) synthesized by the gluconeogenic pathway, and to a lesser extent by the phosphorylation of glycerol. The addition of two fatty acids to glycerol-3-phosphate yields phosphatidate. First, acyl coenzyme A contributes a fatty acid chain to form *lysophosphatidate* and*,* then, a second acyl CoA contributes a fatty acid chain to yield phosphatidate.


R₁CO—CoA
CoA
R₂CO—CoA
CoA
Usually saturated
Usually unsaturated
Glycerol 3-phosphate
Lysophosphatidate
Phosphatidate

These acylations are catalyzed by *glycerol phosphate acyltransferase.* In most phosphatidates, the fatty acid chain attached to the C-1 atom is saturated, whereas the one attached to the C-2 atom is unsaturated. Phosphatidate can also be synthesized from *diacylglycerol* (DAG), in what is essentially a salvage pathway, by the action of *diacylglycerol kinase*:

$$\text{Diacylglycerol} + \text{ATP} \longrightarrow \text{phosphatidate} + \text{ADP}$$

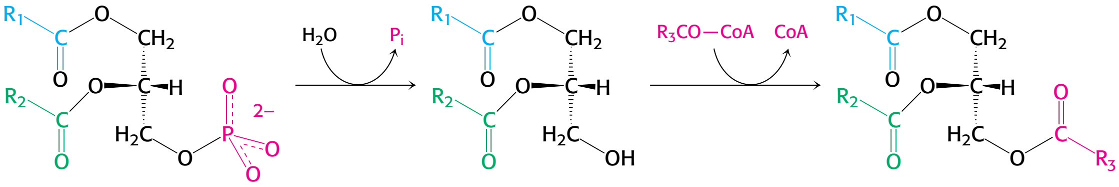
The phospholipid and triacylglycerol pathways diverge at phosphatidate. In the synthesis of triacylglycerols, a key enzyme in the regulation of lipid synthesis, *phosphatidic acid phosphatase*, hydrolyzes phosphatidate to give a diacylglycerol. This intermediate is acylated to a *triacylglycerol* through the addition of a third fatty acid chain in a reaction that is catalyzed by *diglyceride acyltransferase.* Both enzymes are associated in a *triacylglycerol synthetase complex* that is bound to the endoplasmic reticulum membrane.


Lactate
①
DHAP
Glycerol 3-phosphate
Endoplasmic reticulum (all cells)
Phosphatidate*
Alcohol*
LIVER
③
④
Triacylglycerol
Phospholipid
ADIPOSE TISSUE OR DIET
Glycerol
Pᵢ
Triacylglycerol
②
Free fatty acids

*For phospholipid synthesis, either phosphatidate or the alcohol must be activated by reaction with an NTP ④.

Active pathways:
1. Gluconeogenesis, Chapter 16
2. Triacylglycerol breakdown, Chapter 22
3. Triacylglycerol synthesis, Chapter 26
4. Phospholipid synthesis, Chapter 26

FIGURE 26.1 PATHWAY INTEGRATION: Sources of intermediates in the synthesis of triacylglycerols and phospholipids. Phosphatidate, synthesized from dihydroxyacetone phosphate (DHAP) produced in gluconeogenesis and fatty acids, can be further processed to produce triacylglycerol or phospholipids. Phospholipids and other membrane lipids are continuously produced in all cells.



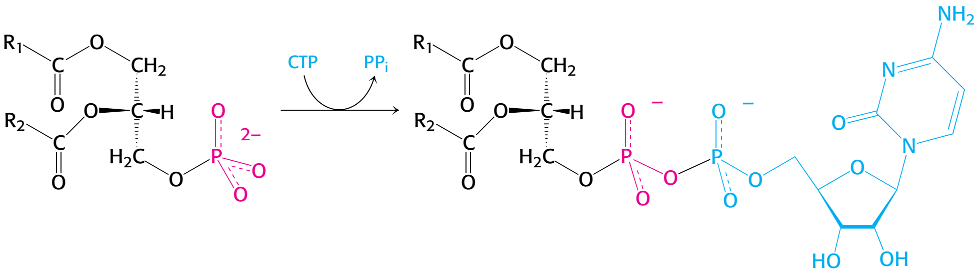
Phosphatidate **Diacylglycerol (DAG)** **Triacylglycerol**

The liver is the primary site of triacylglycerol synthesis. From the liver, the triacylglycerols are transported to the muscles for energy conversion or to the adipose cells for storage.

The synthesis of phospholipids requires an activated intermediate

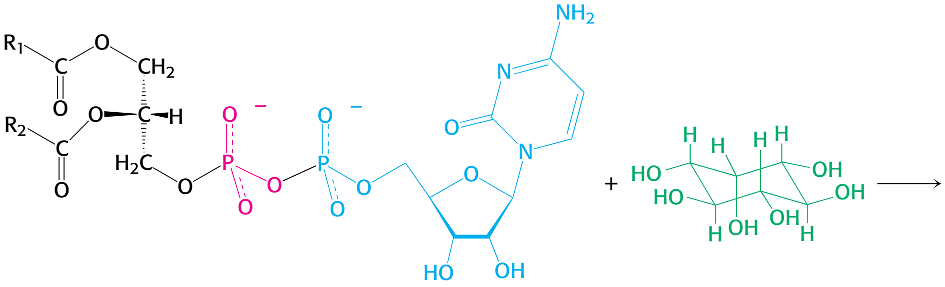
Membrane-lipid synthesis continues in the endoplasmic reticulum and in the Golgi apparatus. *Phospholipid* synthesis requires the combination of a diacylglycerol with an alcohol. As in most anabolic reactions, one of the components must be activated. In this case, either the diacylglycerol or the alcohol may be activated, depending on the source of the reactants.

The synthesis of some phospholipids begins with the reaction of phosphatidate with cytidine triphosphate (CTP) to form the activated diacylglycerol, *cytidine diphosphodiacylglycerol* (CDP-diacylglycerol). This reaction, like those of many biosyntheses, is driven forward by the hydrolysis of pyrophosphate.

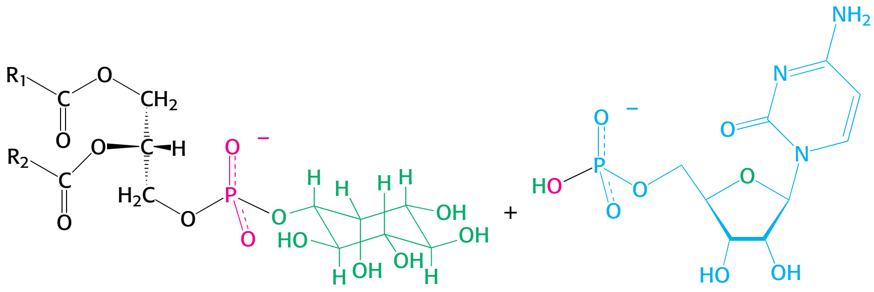


Phosphatidate **CDP-diacylglycerol**

The activated phosphatidyl unit then reacts with the hydroxyl group of an alcohol to form a phosphodiester linkage. If the alcohol is inositol, the products are *phosphatidylinositol* and cytidine monophosphate (CMP).

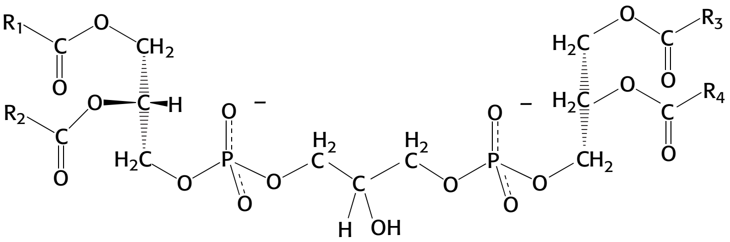


CDP-diacylglycerol **Inositol**



Phosphatidylinositol **CMP**

Subsequent phosphorylations catalyzed by specific kinases lead to the synthesis of *phosphatidylinositol 4,5-bisphosphate,* the precursor of two intracellular messengers—diacylglycerol and inositol 1,4,5-trisphosphate (Section 14.2). If the alcohol is phosphatidylglycerol, the products are *diphosphatidylglycerol (cardiolipin)* and CMP. In eukaryotes, cardiolipin is synthesized in the mitochondria and is located exclusively in inner mitochondrial membranes. Cardiolipin plays an important role in the organization of the protein components of oxidative phosphorylation. For example, it is required for the full activity of cytochrome *c* oxidase (Section 18.3).

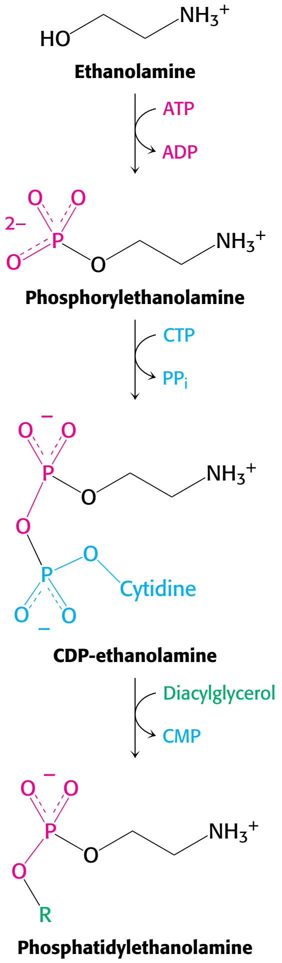


Diphosphatidylglycerol (Cardiolipin)

The fatty acid components of phospholipids may vary, and thus cardiolipin, as well as most other phospholipids, represents a class of molecules rather than a single species. As a result, a single mammalian cell may contain thousands of distinct phospholipids. Phosphatidylinositol is unusual in that it has a nearly fixed fatty acid composition. Stearic acid usually occupies the C-1 position and arachidonic acid the C-2 position.

Some phospholipids are synthesized from an activated alcohol

Phosphatidylethanolamine, the major phospholipid of the inner leaflet of cell membranes, is synthesized from the alcohol ethanolamine. To activate the alcohol, ethanolamine is phosphorylated by ATP to form the precursor, *phosphorylethanolamine.* This precursor then reacts with CTP to form the activated alcohol, *CDP-ethanolamine.* The phosphorylethanolamine unit of CDP-ethanolamine is transferred to a diacylglycerol to form *phosphatidylethanolamine.*

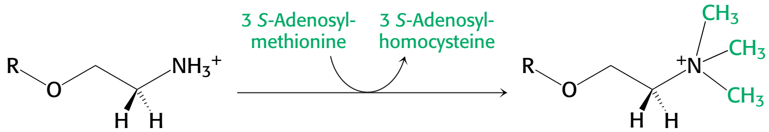


Phosphatidylcholine is an abundant phospholipid

The most common phospholipid in mammals is phosphatidylcholine, comprising approximately 50% of the membrane mass. Dietary choline is activated in a series of reactions analogous to those in the activation of ethanolamine. CTP-phosphocholine cytidylyltransferase (CCT) catalyzes the formation of CDP-choline, the rate-limiting step in phosphatidylcholine synthesis. CCT is an amphitropic enzyme, a class of enzymes whose regulator ligand is the membrane itself. A portion of the enzyme, normally associated with the membrane, detects a fall in phosphatidylcholine as an alteration in the physical properties of the membrane. When this occurs, another portion of the enzyme is inserted into the membrane, leading to enzyme activation. Indeed, the k_{cat}/K_M value (Section 8.4) increases three orders of magnitude upon activation, resulting in the restoration of phosphatidylcholine levels.

Earlier (Section 22.4) we examined how cancer cells increase lipogenesis to meet the fatty acid needs for membrane synthesis. Evidence is accumulating that CCT is specifically activated in some cancers to generate the required phosphocholine.

The importance of phosphatidylcholine is attested to by the fact that the liver possesses an enzyme, *phosphatidylethanolamine methyltransferase,* which synthesizes phosphatidylcholine from phosphatidylethanolamine when dietary choline is insufficient. The amino group of this phosphatidylethanolamine is methylated three times to form *phosphatidylcholine.* S-*Adenosylmethionine* is the methyl donor (Section 24.2).

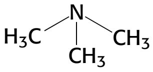


Phosphatidyl-ethanolamine

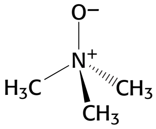
Phosphatidyl-choline

Thus, phosphatidylcholine can be produced by two distinct pathways in mammals, ensuring that this phospholipid can be synthesized even if the components for one pathway are in limited supply.

Excess choline is implicated in the development of heart disease



Trimethylamine



Trimethylamine-*N*-oxide

Choline is a popular dietary supplement that is believed by some to enhance liver and neuronal function. Although the effectiveness of choline supplements is not established, the dangers of excess choline consumption are becoming clear. Gut bacteria convert excess choline into trimethylamine (TMA), a gas that smells like rotten fish, and the liver converts the absorbed TMA into trimethylamine-*N*-oxide (TMAO). TMAO stimulates cholesterol uptake by macrophages, a process that can result in atherosclerosis (p. 787). Foods rich in phosphatidylcholine, such as red meats and dairy products, may also result in TMAO production.

Base-exchange reactions can generate phospholipids

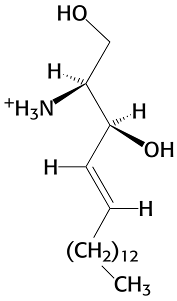
Phosphatidylserine makes up 10% of the phospholipids in mammals. This phospholipid is synthesized in a *base-exchange reaction* of serine with phosphatidylcholine or phosphatidylethanolamine. In the reaction, serine replaces choline or ethanolamine.

$$\text{Phosphatidylcholine} + \text{serine} \longrightarrow \text{choline} + \text{phosphatidylserine}$$

$$\text{Phosphatidylethanolamine} + \text{serine} \longrightarrow \text{ethanolamine} + \text{phosphatidylserine}$$

Phosphatidylserine is normally located in the inner leaflet of the plasma membrane bilayer but is moved to the outer leaflet in apoptosis (Section 18.6). There, it serves to attract phagocytes to consume the cell remnants after apoptosis is complete. Phosphatidylserine is translocated from one side of the membrane to the other by an ATP-binding cassette translocase (Section 13.2).

Note that a cytidine nucleotide plays the same role in the synthesis of these phosphoglycerides as a uridine nucleotide does in the formation of glycogen (Section 21.4). In all of these biosyntheses, an activated intermediate (UDP-glucose, CDP-diacylglycerol, or CDP-alcohol) is formed from a phosphorylated substrate (glucose 1-phosphate, phosphatidate, or a phosphorylalcohol) and a nucleoside triphosphate (UTP or CTP). The activated intermediate then reacts with a hydroxyl group (the terminus of glycogen, an alcohol, or a diacylglycerol).



Sphingosine

Sphingolipids are synthesized from ceramide

We now turn from glycerol-based phospholipids to another class of membrane lipid—the *sphingolipids.* These lipids are found in the plasma membranes of all eukaryotic cells, although the concentration is highest in the cells of the central nervous system. The backbone of a sphingolipid is *sphingosine,* rather than glycerol. Palmitoyl CoA and serine condense to form 3-ketosphinganine. The serine–palmitoyl transferase catalyzing this reaction is the rate-limiting step in the pathway and requires pyridoxal phosphate, revealing again the dominant role of this cofactor in transformations that include amino acids. Ketosphinganine is then reduced to dihydrosphingosine before conversion into *ceramide,* a lipid consisting of a fatty acid chain attached to the amino group of a sphingosine backbone (Figure 26.2).

In all sphingolipids, the amino group of ceramide is acylated. The terminal hydroxyl group also is substituted (Figure 26.3). In *sphingomyelin,* a component of the myelin sheath covering many nerve fibers, the substituent is phosphorylcholine, which comes from phosphatidylcholine. In a *cerebroside,* the substituent is glucose or galactose. UDP-glucose or UDP-galactose is the sugar donor.

Gangliosides are carbohydrate-rich sphingolipids that contain acidic sugars

Gangliosides are the most complex sphingolipids. In a *ganglioside,* an *oligosaccharide chain* is linked to the terminal hydroxyl group of ceramide by a glucose residue (Figure 26.4). This oligosaccharide chain contains at least one acidic sugar, either N-*acetylneuraminate* or N-*glycolylneuraminate.* These acidic sugars are called *sialic acids.* Their nine-carbon backbones are synthesized from phosphoenolpyruvate (a three-carbon unit) and *N*-acetylmannosamine 6-phosphate (a six-carbon unit).

Gangliosides are synthesized by the ordered, step-by-step addition of sugar residues to ceramide. The synthesis of these complex lipids requires

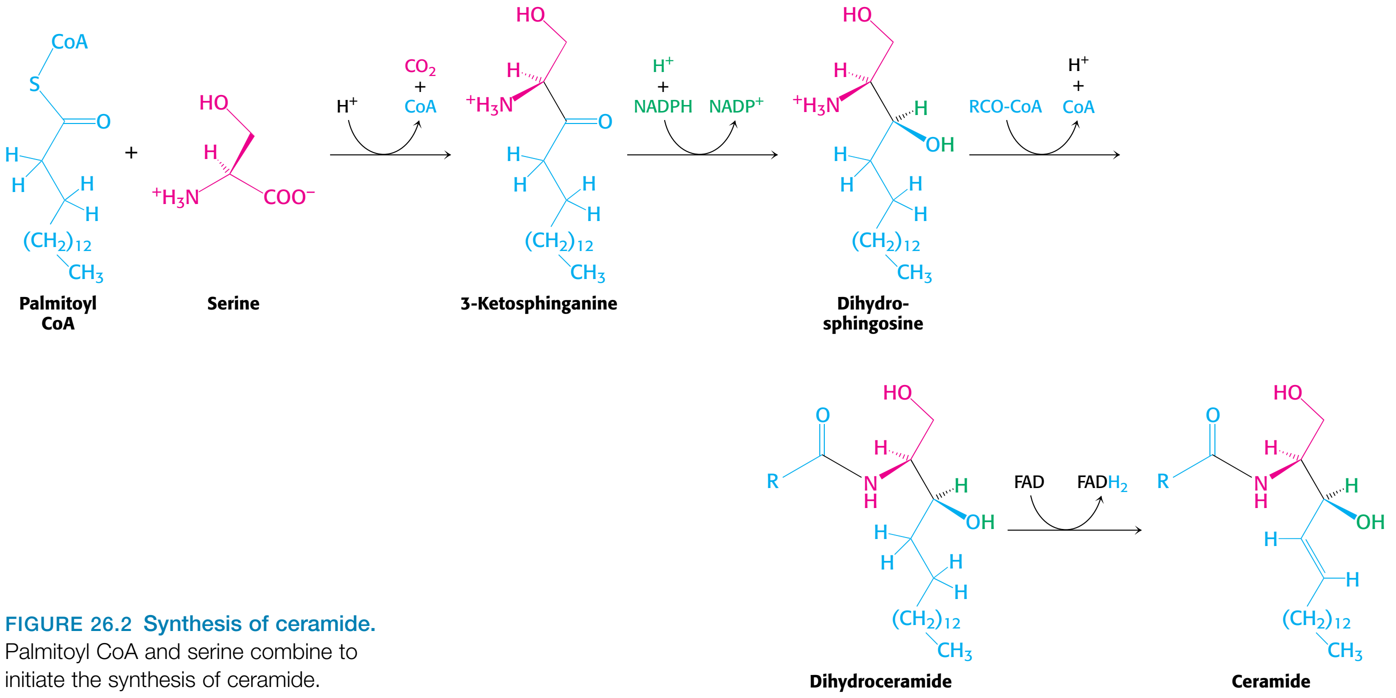


FIGURE 26.2 Synthesis of ceramide. Palmitoyl CoA and serine combine to initiate the synthesis of ceramide.

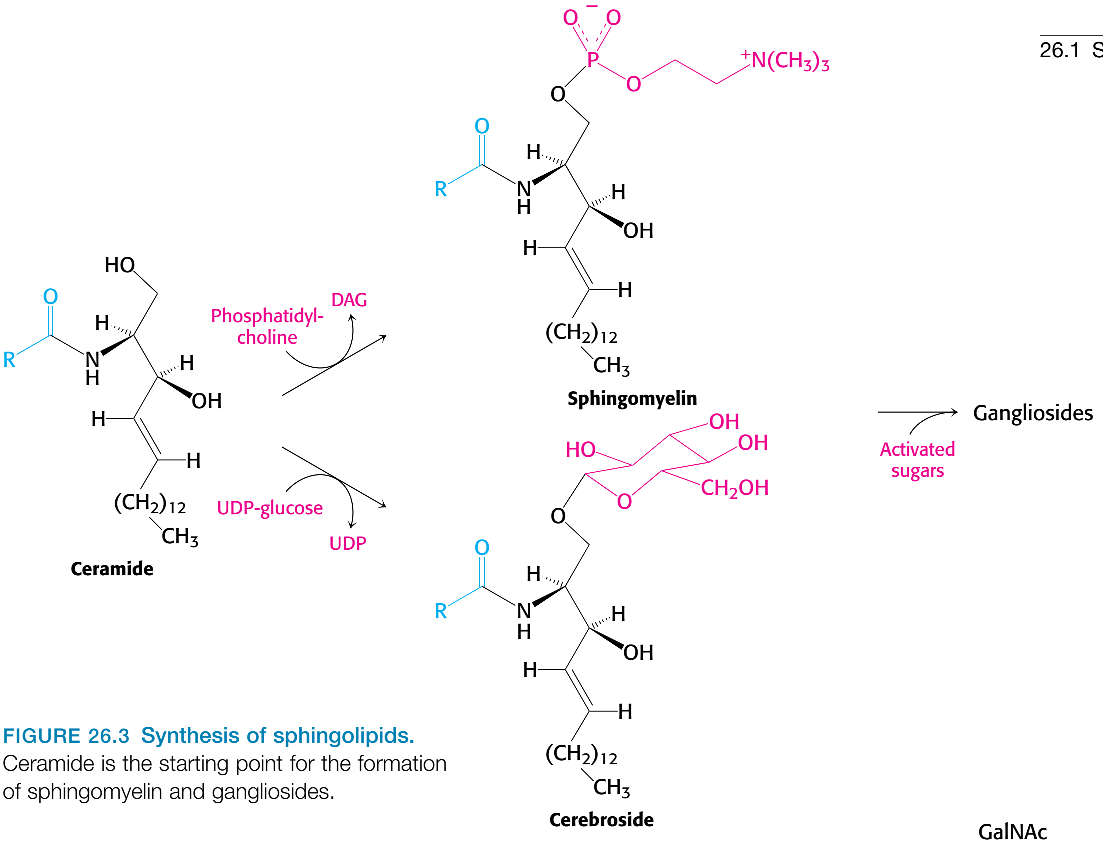


FIGURE 26.3 Synthesis of sphingolipids. Ceramide is the starting point for the formation of sphingomyelin and gangliosides.

the activated sugars UDP-glucose, UDP-galactose, and UDP-*N*-acetylgalactosamine, as well as the CMP derivative of *N*-acetylneuraminate. CMP-*N*-acetylneuraminate is synthesized from CTP and *N*-acetylneuraminate. The sugar composition of the resulting ganglioside is determined by the specificity of the glycosyltransferases in the cell. Almost 200 different gangliosides have been characterized (see Figure 26.4 for the composition of ganglioside G_{M1}).

Ganglioside-binding by the cholera toxin is the first step in the development of cholera, a pathological condition characterized by severe diarrhea (Section 14.5). Enterotoxigenic *E. coli*, the most common cause of diarrhea, including traveler's diarrhea, produces a toxin that also gains access to the cell by first binding to gangliosides. Gangliosides are also crucial for binding immune-system cells to sites of injury in the inflammatory response.

Sphingolipids confer diversity on lipid structure and function

The structures of sphingolipids and the more abundant glycerophospholipids are very similar (Figure 12.8). Given the structural similarity of these two types of lipids, why are sphingolipids required at all? Indeed, the prefix "sphingo" was applied to capture the "sphinxlike" properties of this enigmatic class of lipids. Although the precise role of sphingolipids is not firmly established, progress toward solving the riddle of their function is being made. As discussed in Chapter 12, sphingolipids are important components of lipid rafts, highly organized regions of the plasma membrane that are important in signal transduction. Sphingosine, sphingosine 1-phosphate, and ceramide serve as second messengers in the regulation of cell growth, differentiation, and death. For instance, ceramide

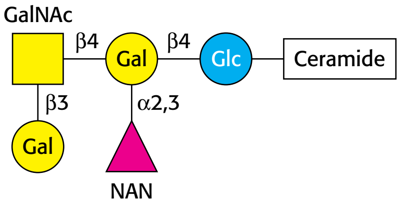
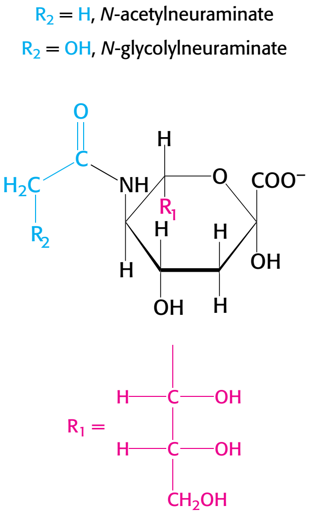


FIGURE 26.4 Ganglioside G_{M1}. This ganglioside consists of five monosaccharides linked to ceramide: one glucose (Glc) molecule, two galactose (Gal) molecules, one *N*-acetylgalactosamine (GalNAc) molecule, and one *N*-acetylneuraminate (NAN) molecule.



derived from a sphingolipid initiates programmed cell death in some cell types (p. 775) and may contribute to the development of type 2 diabetes (Chapter 27).

Respiratory distress syndrome and Tay–Sachs disease result from the disruption of lipid metabolism

Respiratory distress syndrome is a pathological condition resulting from a failure in the biosynthetic pathway for dipalmitoylphosphatidylcholine. This phospholipid, in conjunction with specific proteins and other phospholipids, is found in the extracellular fluid that surrounds the alveoli of the lung. Its function is to decrease the surface tension of the fluid to prevent lung collapse at the end of the expiration phase of breathing. Premature infants may suffer from respiratory distress syndrome because their immature lungs do not synthesize enough dipalmitoylphosphatidylcholine.

Tay–Sachs disease is caused by a failure of lipid degradation: an inability to degrade gangliosides. Gangliosides are found in highest concentration in the nervous system, particularly in gray matter, where they constitute 6% of the lipids. Gangliosides are normally degraded inside lysosomes by the sequential removal of their terminal sugars but, in Tay–Sachs disease, this degradation does not take place. As a consequence, neurons become significantly swollen with lipid-filled lysosomes (Figure 26.5). An affected infant displays weakness and retarded psychomotor skills before 1 year of age. The child is demented and blind by age 2 and usually dies before age 3.

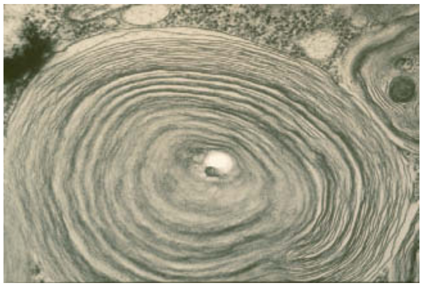
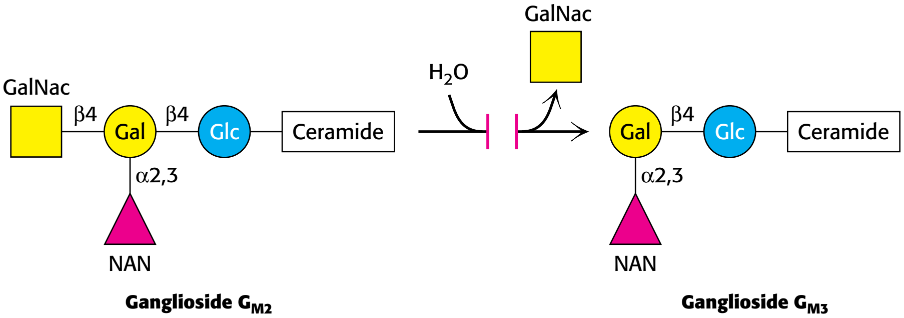


FIGURE 26.5 Lysosome with lipids. An electron micrograph of a lysosome engorged with lipids. Such lysosomes are sometimes described as being "onion-skin" like because the layers of undigested lipids resemble a sliced onion. [Graphics & Photography Service/Custom Medical Stock Photo—All rights reserved.]

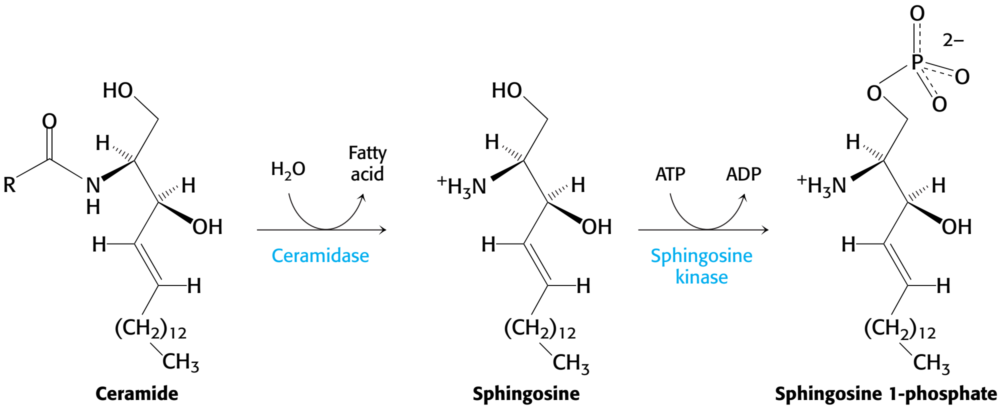
The ganglioside content of the brain of an infant with Tay–Sachs disease is greatly elevated. *The concentration of ganglioside G_{M2} is many times higher than normal because its terminal* N*-acetylgalactosamine residue is removed very slowly or not at all.* The missing or deficient enzyme is a specific *β-N-acetylhexosaminidase.*



Tay–Sachs disease can be diagnosed in the course of fetal development. Cells obtained from either the villi of the placenta (chorionic villus sampling) or from the amniotic fluid (amniocentesis) are tested for the presence of the defective gene. Tay–Sachs disease was especially prominent among Ashkenazi Jews (descendants of Jews from central and eastern Europe). A genetic testing program, initiated in the early 1970s upon the development of a simple blood test to identify carriers, has virtually eliminated the disease in the population.

Ceramide metabolism stimulates tumor growth

Ceramide is a precursor for sphingomyelin, cerebroside, and gangliosides. Ceramide itself, however, induces programmed cell death or apoptosis (Section 18.6). Recall that cancer cells require all types of lipids for membrane formation (p. 771 and Section 22.4). How do cancer cells prevent ceramide-induced cell death? It appears that these cells destroy the apoptotic signal by converting it into a pro-mitotic signal. *Ceramidase* removes the fatty acid from the amino group of ceramide, generating sphingosine. Sphingosine is then converted into sphingosine 1-phosphate by sphingosine kinase, which stimulates cell division.


H₂O
Fatty acid
Ceramidase
ATP
ADP
Sphingosine kinase
Ceramide
Sphingosine
Sphingosine 1-phosphate

Thus, cancer cells convert a potentially lethal signal molecule into one that promotes tumor growth. Efforts are underway to develop inhibitors of ceramidase for use as chemotherapeutic agents.

Phosphatidic acid phosphatase is a key regulatory enzyme in lipid metabolism

Although the details of the regulation of lipid synthesis remain to be elucidated, evidence suggests that *phosphatidic acid phosphatase* (PAP), working in concert with diacylglycerol kinase, plays a key role in the regulation of lipid synthesis. PAP, also called lipin 1 in mammals, controls the extent to which triacylglycerols are synthesized relative to phospholipids and regulates the type of phospholipid synthesized (Figure 26.6). For instance, when PAP activity is high, phosphatidate is dephosphorylated and diacylglycerol is produced, which can react with the appropriate activated alcohols to yield phosphatidylethanolamine, phosphatidylserine, or phosphatidylcholine. Diacylglycerol can also be converted into triacylglycerols, and evidence suggests that the formation of triacylglycerols may act as a fatty acid buffer. This buffering helps to regulate the levels of diacylglycerol and sphingolipids, which serve signaling functions.

When PAP activity is lower, phosphatidate is used as a precursor for different phospholipids, such as phosphatidylinositol and cardiolipin. Moreover, phosphatidate is a signal molecule itself. Phosphatidate

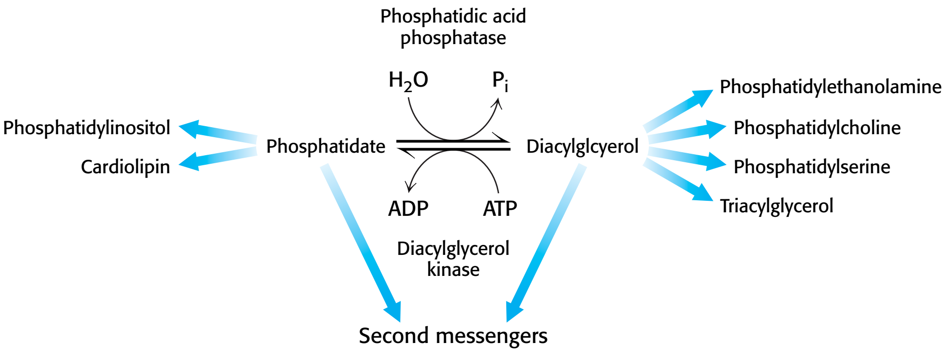

Phosphatidic acid phosphatase
H₂O
Pᵢ
Phosphatidylinositol
Cardiolipin
Phosphatidate
Diacylglcyerol
ADP
ATP
Diacylglycerol kinase
Phosphatidylethanolamine
Phosphatidylcholine
Phosphatidylserine
Triacylglycerol
Second messengers

FIGURE 26.6 Regulation of lipid synthesis. Phosphatidic acid phosphatase is the key regulatory enzyme in lipid synthesis. When active, PAP generates diacylglycerol (DAG), which can react with activated alcohols to form phospholipids or with fatty acyl CoA to form triacylglycerols. When PAP is inactive, phosphatidate is converted into CMP-DAG for the synthesis of different phospholipids. PAP also controls the amount of DAG and phosphatidate, both of which function as second messengers.

regulates the growth of endoplasmic reticulum and nuclear membranes and acts as a cofactor that stimulates the expression of genes in phospholipid synthesis.

What are the signal molecules that regulate the activity of PAP? CDP-diacylglycerol, phosphatidylinositol, and cardiolipin enhance PAP activity, and sphingosine and dihydrosphingosine inhibit it. The enzyme also undergoes extensive phosphorylation and dephosphorylation. When phosphorylated, the enzyme resides in the cytoplasm. Upon dephosphorylation, the enzyme associates with the endoplasmic reticulum, the location of its substrate phosphatidate. Details of the covalent modification are under investigation.

Studies in mice clearly show the importance of PAP for the regulation of fatty acid synthesis. The loss of PAP function prevents normal adipose-tissue development, leading to lipodystrophy (severe loss of body fat) and insulin resistance. Excess PAP activity results in obesity. The regulation of phospholipid synthesis is an exciting area of research that will be active for some time to come.

26.2 Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages

We now turn our attention to the synthesis of the fundamental lipid *cholesterol*. This steroid modulates the fluidity of animal cell membranes (Section 12.5) and is the precursor of steroid hormones such as progesterone, testosterone, estradiol, and cortisol. *All 27 carbon atoms of cholesterol are derived from acetyl CoA* in a three-stage synthetic process (Figure 26.7).

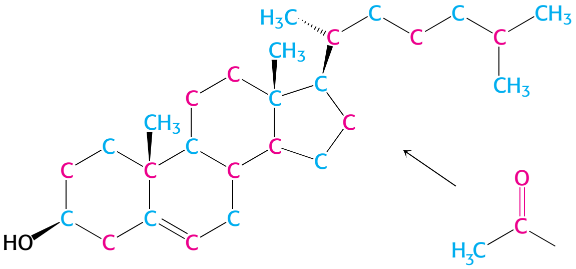


FIGURE 26.7 **Labeling of cholesterol.** Isotope-labeling experiments reveal the source of carbon atoms in cholesterol synthesized from acetate labeled in its methyl group (blue) or carboxylate atom (red).

1. Stage one is the synthesis of isopentenyl pyrophosphate, an activated isoprene unit that is the key building block of cholesterol.
2. Stage two is the condensation of six molecules of isopentenyl pyrophosphate to form squalene.
3. In stage three, squalene cyclizes and the tetracyclic product is subsequently converted into cholesterol.

The first stage takes place in the cytoplasm, and the next two in the endoplasmic reticulum.

Cholesterol

"Cholesterol is the most highly decorated small molecule in biology. Thirteen Nobel Prizes have been awarded to scientists who devoted major parts of their careers to cholesterol. Ever since it was isolated from gallstones in 1784, cholesterol has exerted an almost hypnotic fascination for scientists from the most diverse areas of science and medicine. . . . Cholesterol is a Janus-faced molecule. The very property that makes it useful in cell membranes, namely its absolute insolubility in water, also makes it lethal."

—Michael Brown and Joseph Goldstein, on the occasion of their receipt of a Nobel Prize for elucidating the control of blood levels of cholesterol.
Nobel Lectures (1985) © The Nobel Foundation, 1985

The synthesis of mevalonate, which is activated as isopentenyl pyrophosphate, initiates the synthesis of cholesterol

The first stage in the synthesis of cholesterol is the formation of isopentenyl pyrophosphate from acetyl CoA. This set of reactions starts with the formation of 3-hydroxy-3-methylglutaryl CoA (HMG CoA) from acetyl CoA and acetoacetyl CoA. This intermediate is reduced to *mevalonate* for the synthesis of cholesterol (Figure 26.8). Recall that, alternatively, 3-hydroxy-3-methylglutaryl CoA may be generated in the mitochondria and processed to form ketone bodies, which are subsequently secreted to provide fuel for other tissues, notably the brain under starvation conditions (Section 22.3).

The synthesis of mevalonate is the committed step in cholesterol formation. The enzyme catalyzing this irreversible step, *3-hydroxy-3-methylglutaryl*

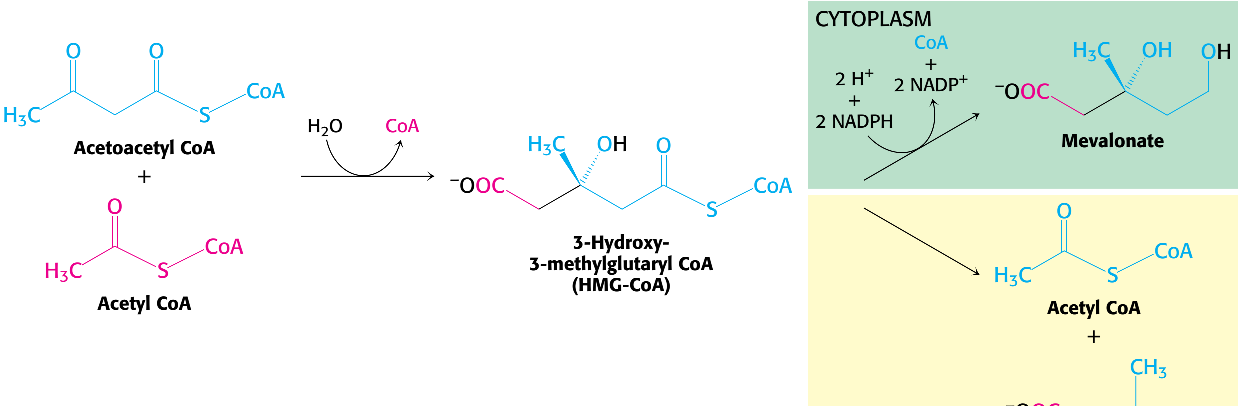

Acetoacetyl CoA
+
Acetyl CoA
H₂O
CoA
3-Hydroxy-3-methylglutaryl CoA (HMG-CoA)
CYTOPLASM
2 H⁺ + 2 NADPH
CoA + 2 NADP⁺
Mevalonate
MITOCHONDRIA
Acetyl CoA
+
Acetoacetate

FIGURE 26.8 Fates of 3-hydroxy-3-methylglutaryl CoA. In the cytoplasm, HMG-CoA is converted into mevalonate. In mitochondria, it is converted into acetyl CoA and acetoacetate.

CoA reductase (HMG-CoA reductase), is an important control site in cholesterol biosynthesis, as will be discussed shortly.

$$\text{3-Hydroxy-3-methylglutaryl CoA} + 2\,\text{NADPH} + 2\,\text{H}^+ \longrightarrow \text{mevalonate} + 2\,\text{NADP}^+ + \text{CoA}$$

HMG-CoA reductase is an integral membrane protein in the endoplasmic reticulum.

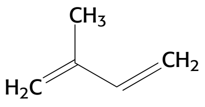
Mevalonate is converted into *3-isopentenyl pyrophosphate* in three consecutive reactions requiring ATP (Figure 26.9). In the last step, the release of CO_2 yields isopentenyl pyrophosphate, an activated isoprene unit that is a key building block for many important biomolecules throughout the kingdoms of life.

Squalene (C_{30}) is synthesized from six molecules of isopentenyl pyrophosphate (C_5)

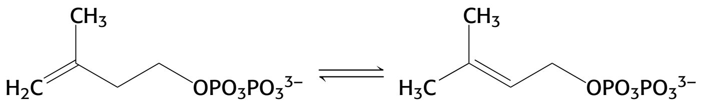
Squalene is synthesized from isopentenyl pyrophosphate by the reaction sequence

$$C_5 \longrightarrow C_{10} \longrightarrow C_{15} \longrightarrow C_{30}$$

This stage in the synthesis of cholesterol starts with the isomerization of *isopentenyl pyrophosphate* to *dimethylallyl pyrophosphate.*



Isoprene


CH₃
H₂C
OPO₃PO₃³⁻
CH₃
H₃C
OPO₃PO₃³⁻

Isopentenyl pyrophosphate **Dimethylallyl pyrophosphate**

These two isomeric C_5 units (one of each type) condense to form a C_{10} compound: isopentenyl pyrophosphate attacks an allylic carbocation

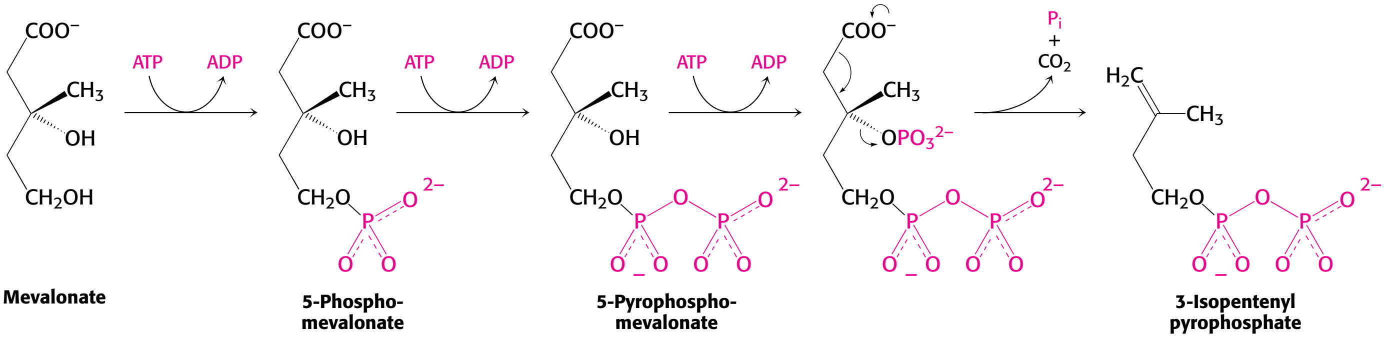

ATP
ADP
ATP
ADP
ATP
ADP
Pᵢ + CO₂
Mevalonate
5-Phospho-mevalonate
5-Pyrophospho-mevalonate
3-Isopentenyl pyrophosphate

FIGURE 26.9 Synthesis of isopentenyl pyrophosphate. This activated intermediate is formed from mevalonate in three steps requiring ATP, followed by a decarboxylation.

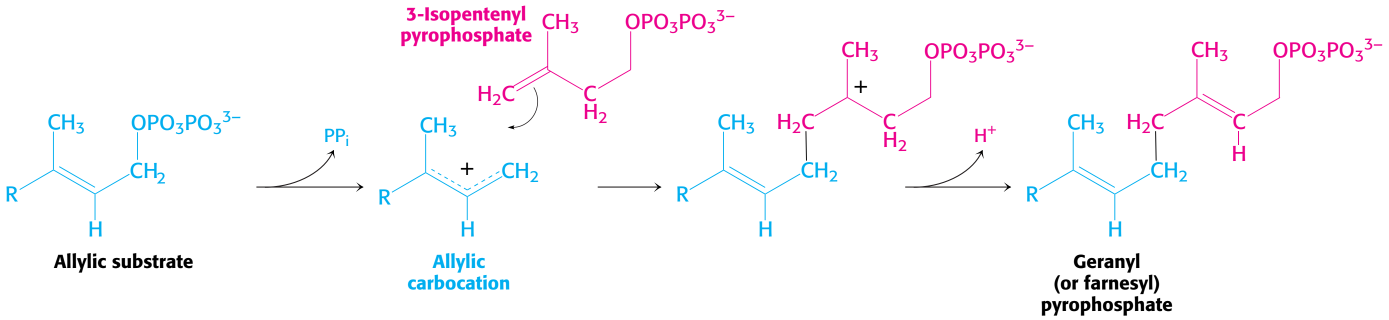

3-Isopentenyl pyrophosphate
CH$_3$
OPO$_3$PO$_3$$^{3-}$
H$_2$C
C H$_2$
PP$_i$
H$^+$
R
H
CH$_2$
+
Allylic substrate
Allylic carbocation
Geranyl (or farnesyl) pyrophosphate

FIGURE 26.10 **Condensation mechanism in cholesterol synthesis.** The isomeric C_5 units dimethylallyl pyrophosphate and isopentenyl pyrophosphate join to form geranyl pyrophosphate. The same mechanism is used to add an additional isopentenyl pyrophosphate to form farnesyl pyrophosphate.

formed from dimethylallyl pyrophosphate to yield *geranyl pyrophosphate* (Figure 26.10). The same kind of reaction takes place again: geranyl pyrophosphate is converted into an allylic carbonium ion, which is attacked by isopentenyl pyrophosphate. The resulting C_{15} compound is called *farnesyl pyrophosphate.* The same enzyme, *geranyl transferase,* catalyzes each of these condensations.

The last step in the synthesis of *squalene* is a reductive tail-to-tail condensation of two molecules of farnesyl pyrophosphate catalyzed by the endoplasmic reticulum enzyme *squalene synthase.*

$$\begin{aligned}&2\text{ Farnesyl pyrophosphate}(C_{15})\\&\quad + \text{NADPH} \longrightarrow \text{squalene}(C_{30})\\&\qquad\qquad + 2\,PP_i + NADP^+ + H^+\end{aligned}$$

The reactions leading from C_5 units to squalene, a C_{30} isoprenoid, are summarized in Figure 26.11.

Squalene cyclizes to form cholesterol

The final stage of cholesterol biosynthesis starts with the cyclization of squalene (Figure 26.12). Squalene is first activated by conversion into squalene epoxide (2,3-oxidosqualene) in a reaction that uses O_2 and NADPH. Squalene epoxide is then cyclized to *lanosterol* by *oxidosqualene cyclase*. The enzyme holds squalene epoxide in an appropriate conformation and initiates

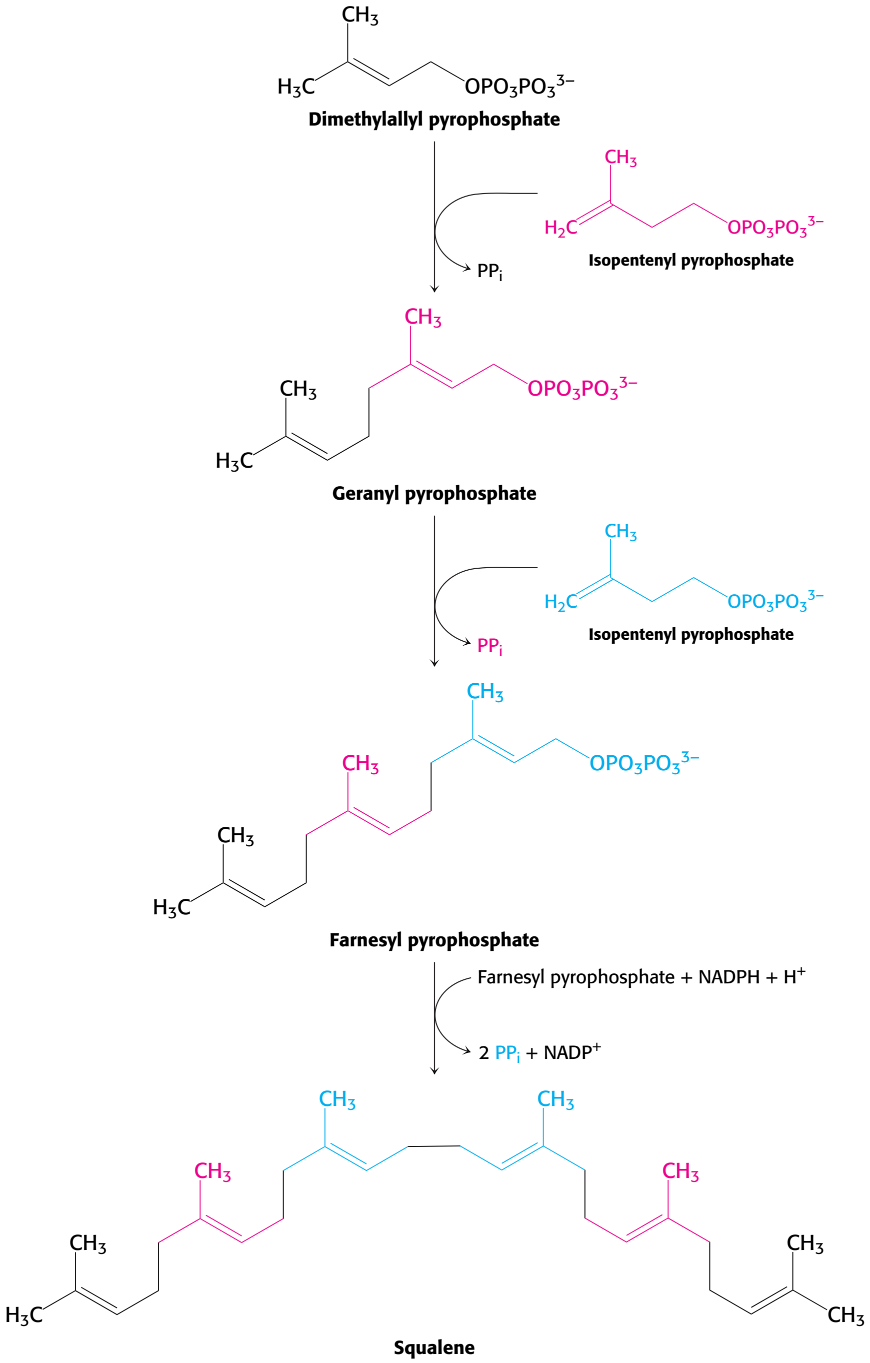

CH$_3$
H$_3$C
OPO$_3$PO$_3$$^{3-}$
Dimethylallyl pyrophosphate
H$_2$C
Isopentenyl pyrophosphate
PP$_i$
Geranyl pyrophosphate
Isopentenyl pyrophosphate
PP$_i$
Farnesyl pyrophosphate
Farnesyl pyrophosphate + NADPH + H$^+$
2 PP$_i$ + NADP$^+$
Squalene

FIGURE 26.11 **Squalene synthesis.** One molecule of dimethyallyl pyrophosphate and two molecules of isopentenyl pyrophosphate condense to form farnesyl pyrophosphate. The tail-to-tail coupling of two molecules of farnesyl pyrophosphate yields squalene.

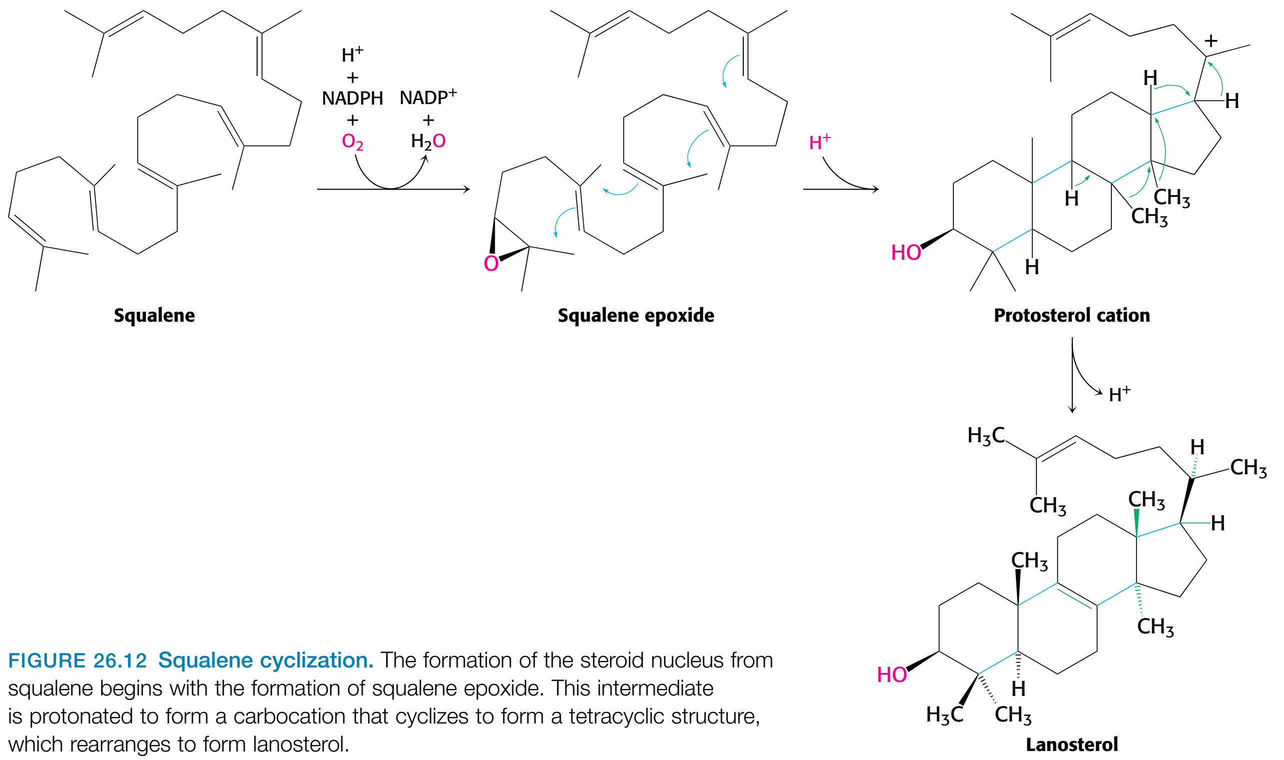


FIGURE 26.12 Squalene cyclization. The formation of the steroid nucleus from squalene begins with the formation of squalene epoxide. This intermediate is protonated to form a carbocation that cyclizes to form a tetracyclic structure, which rearranges to form lanosterol.

the reaction by protonating the epoxide oxygen. The carbocation formed spontaneously rearranges to produce lanosterol. Lanosterol is converted into cholesterol in a multistep process by the removal of three methyl groups, the reduction of one double bond by NADPH, and the migration of the other double bond (Figure 26.13).

26.3 The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels

Cholesterol can be obtained from the diet or it can be synthesized de novo. Cholesterol biosynthesis is one of the most highly regulated metabolic pathways known. Biosynthetic rates may vary several hundredfold, depending on how much cholesterol is consumed in the diet. An adult on a low-cholesterol diet typically synthesizes about 800 mg of cholesterol per day. The liver is the major site of cholesterol synthesis in mammals, although the intestine also forms significant amounts. The rate of cholesterol formation by these organs is highly responsive to the cellular level of cholesterol. *This feedback regulation is mediated primarily by changes in the amount and activity of 3-hydroxy-3-methylglutaryl CoA reductase*. As described earlier (p. 776), this enzyme catalyzes the formation of mevalonate, the committed step in cholesterol biosynthesis. HMG CoA reductase is controlled in multiple ways:

1. The rate of *synthesis of reductase mRNA* is controlled by the *sterol regulatory element binding protein* (SREBP), a member of a family of transcription factors that regulate the proteins required for lipid synthesis. This transcription factor binds to a short DNA sequence called the *sterol regulatory element* (SRE) on the 5′ side of the reductase gene when cholesterol levels are low, enhancing transcription of the gene. In its inactive state, the SREBP resides in the endoplasmic reticulum membrane, where it is associated with

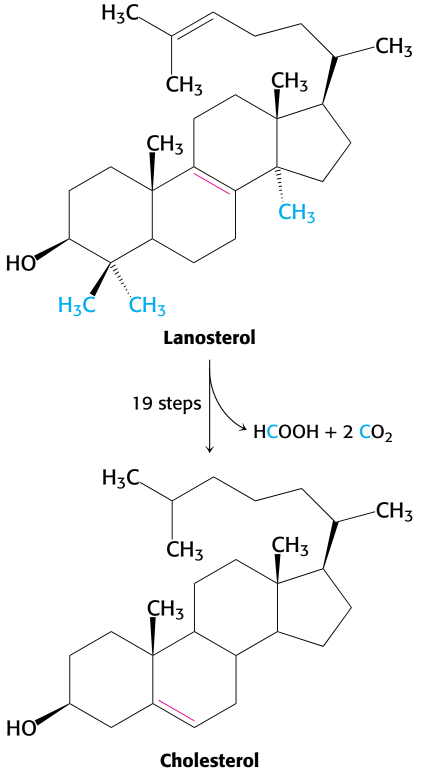


FIGURE 26.13 Cholesterol formation. Lanosterol is converted into cholesterol in a complex process.

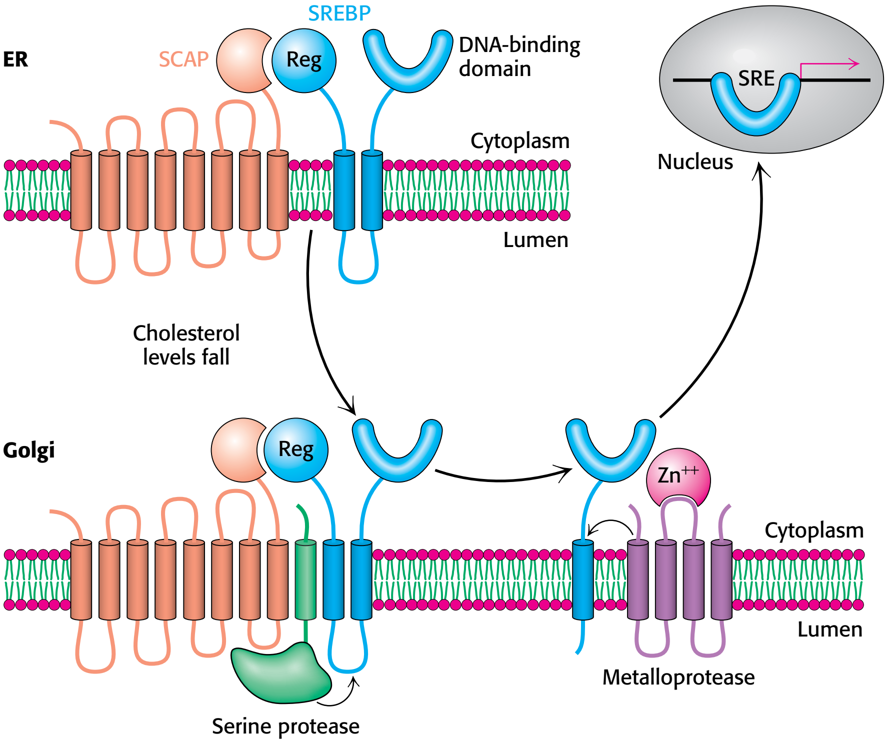


FIGURE 26.14 The SREBP pathway. SREBP resides in the endoplasmic reticulum (ER), where it is bound to SCAP by its regulatory (Reg) domain. When cholesterol levels fall, SCAP and SREBP move to the Golgi complex, where SREBP undergoes successive proteolytic cleavages by a serine protease and a metalloprotease. The released DNA-binding domain moves to the nucleus to alter gene expression. [Information from an illustration provided by Dr. Michael Brown and Dr. Joseph Goldstein.]

the *S*REBP *c*leavage *a*ctivating *p*rotein (SCAP), an integral membrane protein. SCAP is the cholesterol sensor. When cholesterol levels fall, SCAP escorts SREBP in small membrane vesicles to the Golgi complex, where it is released from the membrane by two specific proteolytic cleavages (Figure 26.14). The first cleavage frees a fragment of SREBP from SCAP, whereas the second cleavage releases the regulatory domain from the membrane. The released protein migrates to the nucleus and binds the SRE of the HMG-CoA reductase gene, as well as several other genes in the cholesterol biosynthetic pathway, to enhance transcription. When cholesterol levels rise, the proteolytic release of the SREBP is blocked, and the SREBP in the nucleus is rapidly degraded by proteasomes located in the nucleus. These two events halt the transcription of genes of the cholesterol biosynthetic pathways.

What is the molecular mechanism that retains SCAP–SREBP in the ER when cholesterol is present but allows movement to the Golgi complex when cholesterol concentration is low? When cholesterol is low, SCAP binds to vesicular proteins that facilitate the transport of SCAP–SREBP to the Golgi apparatus, as heretofore described. When cholesterol is present, SCAP binds cholesterol, which causes a structural change in SCAP so that it binds to another endoplasmic reticulum protein called Insig (*ins*ulin-*i*nduced *g*ene) (Figure 26.15). Insig is the anchor that retains SCAP and thus SREBP in the endoplasmic reticulum in the presence of cholesterol. The interactions between SCAP and Insig can also be forged when Insig binds 25-hydroxycholesterol, a metabolite of cholesterol. Thus, two distinct steroid–protein interactions prevent the inappropriate movement of SCAP–SREBP to the Golgi complex.

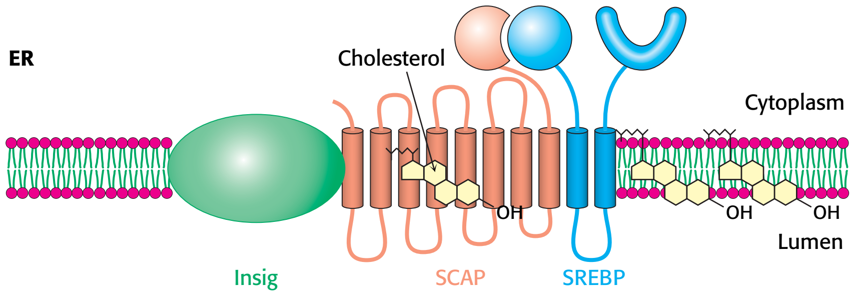


FIGURE 26.15 Insig regulates SCAP–SREBP movement. In the presence of cholesterol, Insig interacts with SCAP–SREBP and prevents the activation of SREBP. Cholesterol binding to SCAP or 25-hydroxycholesterol binding to Insig facilitates the interaction of Insig and SCAP, retaining SCAP–SREBP in the endoplasmic reticulum. [Information from M. S. Brown and J. L. Goldstein. Cholesterol feedback: From Schoenheimer's bottle to Scap's MELADL, *J. Lipid Res.* 50:S15–S27, 2009.]

2. The rate of *translation of reductase mRNA* is inhibited by nonsterol metabolites derived from mevalonate.

3. The *degradation of the reductase* is stringently controlled. The enzyme is bipartite: its cytoplasmic domain carries out catalysis and *its membrane domain senses signals that lead to its degradation.* The membrane domain may undergo structural changes in *response to increasing concentrations of sterols such as lanosterol and 25-hydroxycholesterol.* Under these conditions, the reductase appears to bind to a subset of Insigs that are also associated with the ubiquitinating enzymes (Figure 26.16). The reductase is polyubiquitinated and subsequently extracted from the membrane in a process that requires gerenylgeraniol. The extracted reductase is then degraded by the proteasome (Section 23.2). The combined regulation at the levels of transcription, translation, and degradation can alter the amount of enzyme in the cell more than 200-fold.

4. *Phosphorylation decreases the activity of the reductase.* This enzyme, like acetyl CoA carboxylase (which catalyzes the committed step in fatty acid synthesis, Section 22.5), is switched off by an AMP-activated protein kinase. Thus, cholesterol synthesis ceases when the ATP level is low.

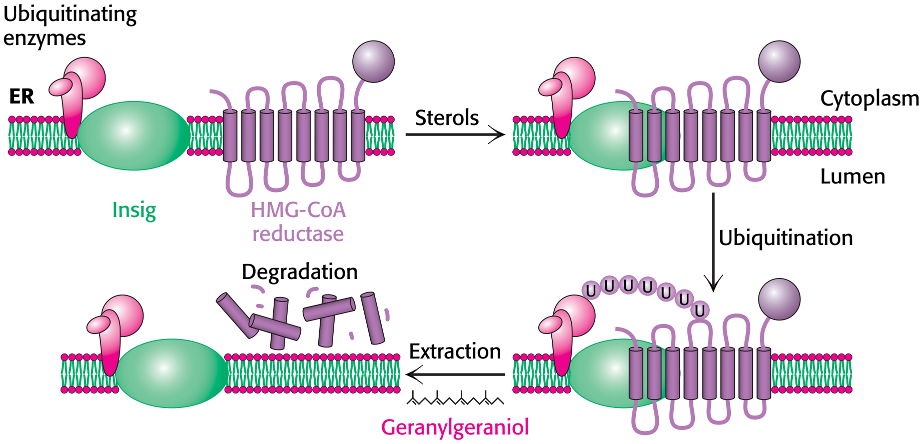


FIGURE 26.16 Insig facilitates the degradation of HMG-CoA reductase. In the presence of sterols, a subclass of Insig associated with ubiquitinating enzymes binds HMG-CoA reductase. This interaction results in the ubiquitination of the enzyme. This modification and the presence of geranylgeraniol results in extraction of the enzyme from the membrane and degradation by the proteasome. [Information from R. A. DeBose-Boyd. Feedback regulation of cholesterol synthesis: Sterol-accelerated ubiquitination and degradation of HMG CoA reductase, *Cell Res.* 18:609–621, 2008.]

TABLE 26.1 Properties of plasma lipoproteins

					Composition (%)				
Plasma lipoproteins	Density (g ml^{-1})	Diameter (nm)	Apolipoprotein	Physiological role	TAG	CE	C	PL	P
Chylomicron	<0.95	75–1200	B-48, C, E	Dietary fat transport	86	3	1	8	2
Very low density lipoprotein	0.95–1.006	30–80	B-100, C, E	Endogenous fat transport	52	14	7	18	8
Intermediate-density lipoprotein	1.006–1.019	15–35	B-100, E	LDL precursor	38	30	8	23	11
Low-density lipoprotein	1.019–1.063	18–25	B-100	Cholesterol transport	10	38	8	22	21
High-density lipoprotein	1.063–1.21	7.5–20	A	Reverse cholesterol transport	5–10	14–21	3–7	19–29	33–57

Abbreviations: TAG, triacylglyerol; CE, cholesteryl ester; C, free cholesterol; PL, phospholipid; P, protein.

Lipoproteins transport cholesterol and triacylglycerols throughout the organism

Cholesterol and triacylglycerols are transported in body fluids in the form of *lipoprotein particles,* which are important for a number of reasons. First, lipoprotein particles are the means by which triacylglycerols are delivered to tissues, from the intestine or liver, for use as fuel or for storage. Second, the fatty acid constituents of the triacylglycerol components of the lipoprotein particles are incorporated into phospholipids for membrane synthesis. Likewise, cholesterol is a vital component of membranes and is a precursor to the powerful signal molecules, the steroid hormones. Finally, cells are not able to degrade the steroid nucleus. Consequently, the cholesterol must be used biochemically or excreted by the liver. Excess cholesterol plays a role in the development of atherosclerosis. Lipoprotein particles function in cholesterol homeostasis, transporting the molecule from sites of synthesis to sites of use, and finally to the liver for excretion.

Each lipoprotein particle consists of a core of hydrophobic lipids surrounded by a shell of more-polar lipids and proteins. The protein components of these macromolecular aggregates, called apoproteins, have two roles: *they solubilize hydrophobic lipids and contain cell-targeting signals.* Apolipoproteins are synthesized and secreted by the liver and the intestine. Lipoprotein particles are classified according to increasing density (Table 26.1): *chylomicrons, chylomicron remnants, very low density lipoproteins* (VLDLs), *intermediate-density lipoproteins* (IDLs), *low-density lipoproteins* (LDLs), and *high-density lipoproteins* (HDLs). These classes have numerous subtypes. Moreover, lipoprotein particles can shift between classes as they release or pick up cargo, thereby changing their density.

Triacylglycerols, cholesterol, and other lipids obtained from the diet are carried away from the intestine in the form of large *chylomicrons* (Section 22.1). These particles have a very low density because triacylglycerols constitute about 90% of their content. Apolipoprotein B-48 (apo B-48), a large protein (240 kDa), forms an amphipathic spherical shell around the fat globule; the external face of this shell is hydrophilic. The triacylglycerols in chylomicrons are released through hydrolysis by *lipoprotein lipases.* These enzymes are located on the lining of blood vessels in muscle and other tissues that use fatty acids as fuels or in the synthesis of lipids. The liver then takes up the cholesterol-rich residues, known as *chylomicron remnants.*

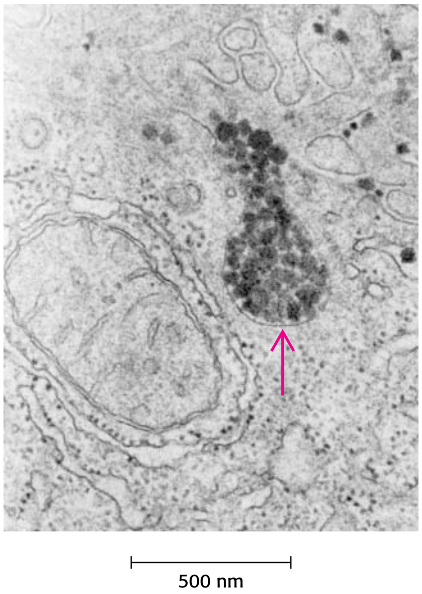


FIGURE 26.17 Site of cholesterol synthesis. Electron micrograph of a part of a liver cell actively engaged in the synthesis and secretion of very low density lipoprotein (VLDL). The arrow points to a vesicle that is releasing its content of VLDL particles. [Courtesy of Dr. George Palade/Yale University, Harvey Cushing/John Hay Whitney Medical Library.]

Lipoprotein particles are also crucial for the transport of lipids from the liver, which is a major site of triacylglycerol and cholesterol synthesis, to other tissues in the body (Figure 26.17). Triacylglycerols and cholesterol in excess of the liver's own needs are exported into the blood in the form of very low density lipoproteins. These particles are stabilized by two apolipoproteins—apo B-100 (513 kDa) and apo E (34 kDa). Triacylglycerols in very low density

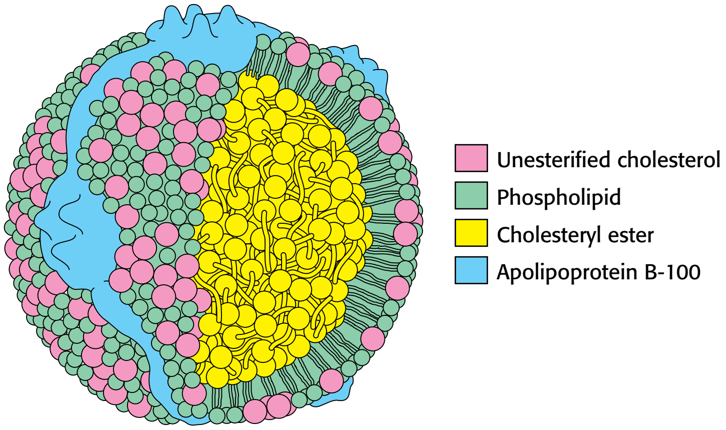

Unesterified cholesterol
Phospholipid
Cholesteryl ester
Apolipoprotein B-100

FIGURE 26.18 Schematic model of low-density lipoprotein. The LDL particle is approximately 22 nm (220 Å) in diameter.

lipoproteins, as in chylomicrons, are hydrolyzed by lipases on capillary surfaces, with the released fatty acids being taken up by the muscle and other tissues. The resulting remnants, which are rich in cholesteryl esters, are called *intermediate-density lipoproteins*. These particles have two fates. Half of them are taken up by the liver for processing, and half are converted into low-density lipoprotein by the removal of more triacylglycerol by tissue lipases that absorb the released fatty acids. Interestingly, apo B-100, one of the largest proteins known, is a longer version of apo B-48, the protein component of chylomicrons. Both apo B proteins are encoded by the same gene and produced from the same initial RNA transcript. In the intestine, RNA editing (Section 29.3) modifies the transcript to generate the mRNA for apo B-48, the truncated form.

Low-density lipoprotein is the major carrier of cholesterol in blood (Figure 26.18). It contains a core of some 1500 cholesterol molecules esterified to fatty acids; the most common fatty acid chain in these esters is linoleate, a polyunsaturated fatty acid. A shell of phospholipids and unesterified cholesterol molecules surrounds this highly hydrophobic core. The shell also contains a single copy of apo B-100, which is recognized by target cells. *The role of LDL is to transport cholesterol to peripheral tissues and regulate de novo cholesterol synthesis at these sites,* as described on page 784. A different purpose is served by *high-density lipoprotein*, which picks up cholesterol released into the plasma from dying cells and from membranes undergoing turnover and delivers the cholesterol to the liver for excretion. An acyltransferase in HDL esterifies these cholesterols, which are then returned by HDL to the liver (Figure 26.19).

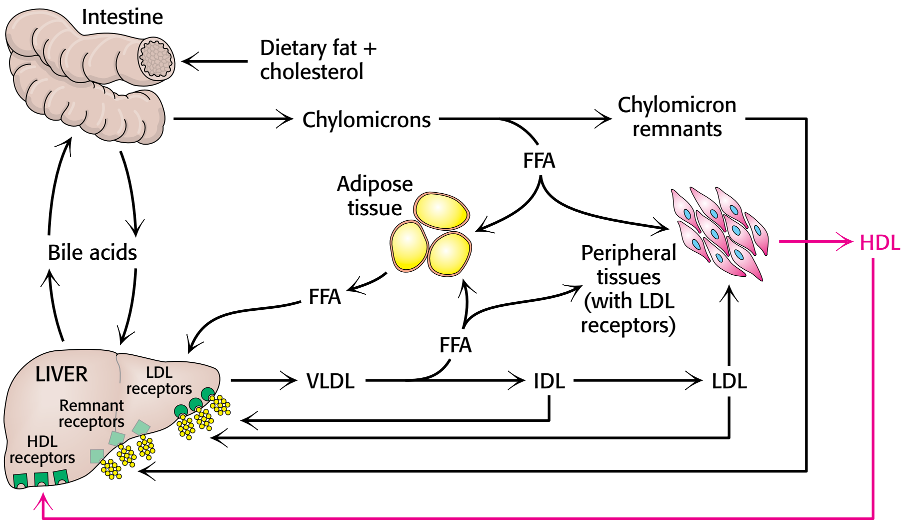

Intestine
Dietary fat + cholesterol
Chylomicrons
Chylomicron remnants
FFA
Adipose tissue
Bile acids
Peripheral tissues (with LDL receptors)
HDL
FFA
FFA
LIVER
LDL receptors
Remnant receptors
HDL receptors
VLDL
IDL
LDL

FIGURE 26.19 An overview of lipoprotein particle metabolism. Fatty acids are abbreviated FFA. [Information from J. G. Hardman (Ed.), L. L. Limbird (Ed.), and A. G. Gilman (Consult. Ed.), *Goodman and Gilman's The Pharmacological Basis of Therapeutics,* 10th ed. (McGraw-Hill, 2001), p. 975, Fig. 36.1.]

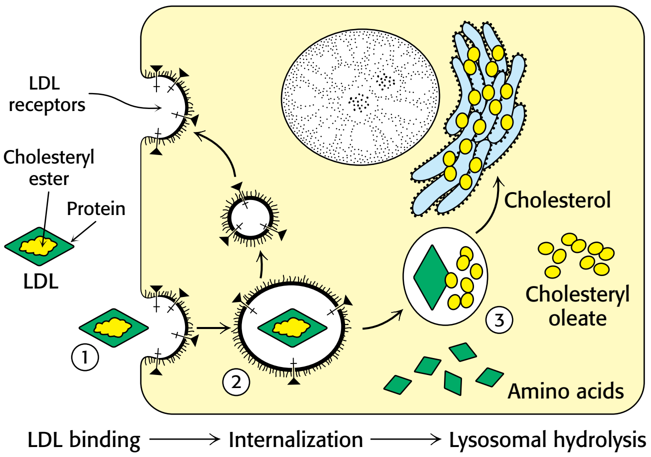


FIGURE 26.20 Receptor-mediated endocytosis. The process of receptor-mediated endocytosis is illustrated for the cholesterol-carrying complex, low-density lipoprotein (LDL): (1) LDL binds to a specific receptor, the LDL receptor; (2) this complex invaginates to form an endosome; (3) after separation from its receptor, the LDL-containing vesicle fuses with a lysosome, leading to the degradation of the LDL and the release of the cholesterol.

Low-density lipoproteins play a central role in cholesterol metabolism

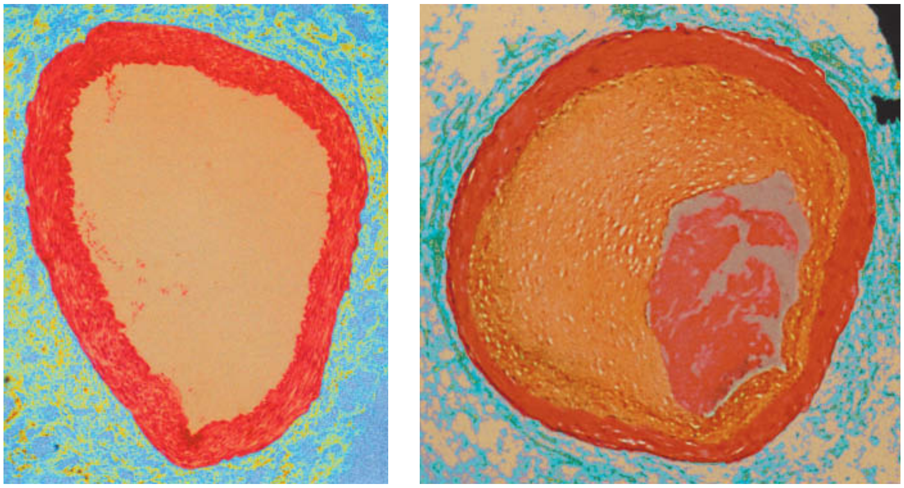
Cholesterol metabolism must be precisely regulated to prevent atherosclerosis. The mode of control in the liver, the primary site of cholesterol synthesis, has already been considered: dietary cholesterol reduces the activity and amount of 3-hydroxy-3-methylglutaryl CoA reductase, the enzyme catalyzing the committed step. In general, cells outside the liver and intestine obtain cholesterol from the plasma rather than synthesizing it de novo. Specifically, *their primary source of cholesterol is the low-density lipoprotein.* The process of LDL uptake, called *receptor-mediated endocytosis,* serves as a paradigm for the uptake of many molecules (Figure 26.20).

Endocytosis begins when apolipoprotein B-100 on the surface of an LDL particle binds to a specific receptor protein on the plasma membrane of nonhepatic cells. The receptors for LDL are localized in specific regions called *coated pits,* which contain a protein called *clathrin.* The receptor–LDL complex is then internalized by *endocytosis;* that is, the plasma membrane in the vicinity of the complex invaginates and then fuses to form an endocytic vesicle called an *endosome*. The endosome is acidified by an ATP-dependent proton pump homologous to the Na^+/K^+ ATPase (Section 13.2), which causes the receptor to release its cargo. Some of the receptor is returned to the cell membrane in a recycling vesicle, while a portion is degraded along with its cargo (p. 787). The round-trip time for a receptor is about 10 minutes; in its lifetime of about a day, each receptor may bring hundreds of LDL particles into the cell. The vesicles containing LDL and in some cases the receptor subsequently fuse with *lysosomes,* acidic vesicles that carry a wide array of degradative enzymes. The protein component of LDL is hydrolyzed to free amino acids. The cholesteryl esters in LDL are hydrolyzed by a lysosomal acid lipase. *The released unesterified cholesterol can then be used for membrane biosynthesis.* Alternatively, it can be *reesterified for storage inside the cell.* In fact, free cholesterol activates *acyl CoA:cholesterol acyltransferase* (ACAT), the enzyme catalyzing this reaction. Reesterified cholesterol contains mainly oleate and palmitoleate, which are monounsaturated fatty acids, in contrast with the cholesteryl esters in LDL, which are rich in linoleate, a polyunsaturated fatty acid (Table 12.1). It is imperative that the cholesterol be reesterified. High concentrations of unesterified cholesterol disrupt the integrity of cell membranes.

The synthesis of the LDL receptor is itself subject to feedback regulation. Studies of cultured fibroblasts show that, *when cholesterol is abundant inside the cell, new LDL receptors are not synthesized, and so the uptake of additional cholesterol from plasma LDL is blocked.* The gene for the LDL receptor, like that for the reductase, is regulated by SREBP, which binds to a sterol regulatory element that controls the rate of mRNA synthesis.

The absence of the LDL receptor leads to hypercholesterolemia and atherosclerosis

The pioneering studies of *familial hypercholesterolemia* by Michael Brown and Joseph Goldstein revealed the physiological importance of the LDL receptor. The total concentration of cholesterol and LDL in the blood plasma is markedly elevated in this genetic disorder, which results from a mutation at a single autosomal locus. The cholesterol level in the plasma of homozygotes is typically 680 mg dl^{-1}, compared with



(A) (B)

FIGURE 26.21 The effects of excess cholesterol. Cross section of (A) a normal artery and (B) an artery blocked by a cholesterol-rich plaque. [SPL/Photo Researchers.]

300 mg dl^{-1} in heterozygotes (clinical assay results are often expressed in milligrams per deciliter, which is equal to milligrams per 100 milliliters). A value of < 200 mg dl^{-1} is regarded as desirable, but many people have higher levels. *In familial hypercholesterolemia, cholesterol is deposited in various tissues because of the high concentration of LDL cholesterol in the plasma.* Nodules of cholesterol called *xanthomas* are prominent in skin and tendons. LDL accumulates under the endothelial cells lining the blood vessels. Of particular concern is the oxidation of the excess LDL to form oxidized LDL (oxLDL), which can instigate the inflammatory response by the immune system, a response that has been implicated in the development of cardiovascular disease. The oxLDL is taken up by immune system cells called macrophages, which become engorged to form foam cells. These foam cells become trapped in the walls of the blood vessels and contribute to the formation of atherosclerotic plaques that cause arterial narrowing and lead to heart attacks (Figure 26.21). In fact, *most homozygotes die of coronary artery disease in childhood.* The disease in heterozygotes (1 in 500 people) has a milder and more variable clinical course.

The molecular defect in most cases of familial hypercholesterolemia is an absence or deficiency of functional receptors for LDL. Receptor mutations that disrupt each of the stages in the endocytotic pathway have been identified. Homozygotes have almost no functional receptors for LDL, whereas heterozygotes have about half the normal number. Consequently, the entry of LDL into liver and other cells is impaired, leading to an increased level of LDL in the blood plasma. Furthermore, less IDL enters liver cells because IDL entry, too, is mediated by the LDL receptor. Consequently, IDL stays in the blood longer in familial hypercholesterolemia, and more of it is converted into LDL than in normal people. All deleterious consequences of an absence or deficiency of the LDL receptor can be attributed to the ensuing elevated level of LDL cholesterol in the blood.

Mutations in the LDL receptor prevent LDL release and result in receptor destruction

One class of mutations that results in familial hypercholesterolemia generates receptors that are reluctant to give up the LDL cargo. Let us begin by examining the makeup of the receptor. The human LDL receptor is a 160-kDa glycoprotein, which is composed of six different types of domains (Figure 26.22A). The amino-terminal region of the receptor, which is the site of LDL binding, consists of seven homologous *LDL receptor type-A* (LA) domains, with domains 4 and 5 most important for LDL binding.

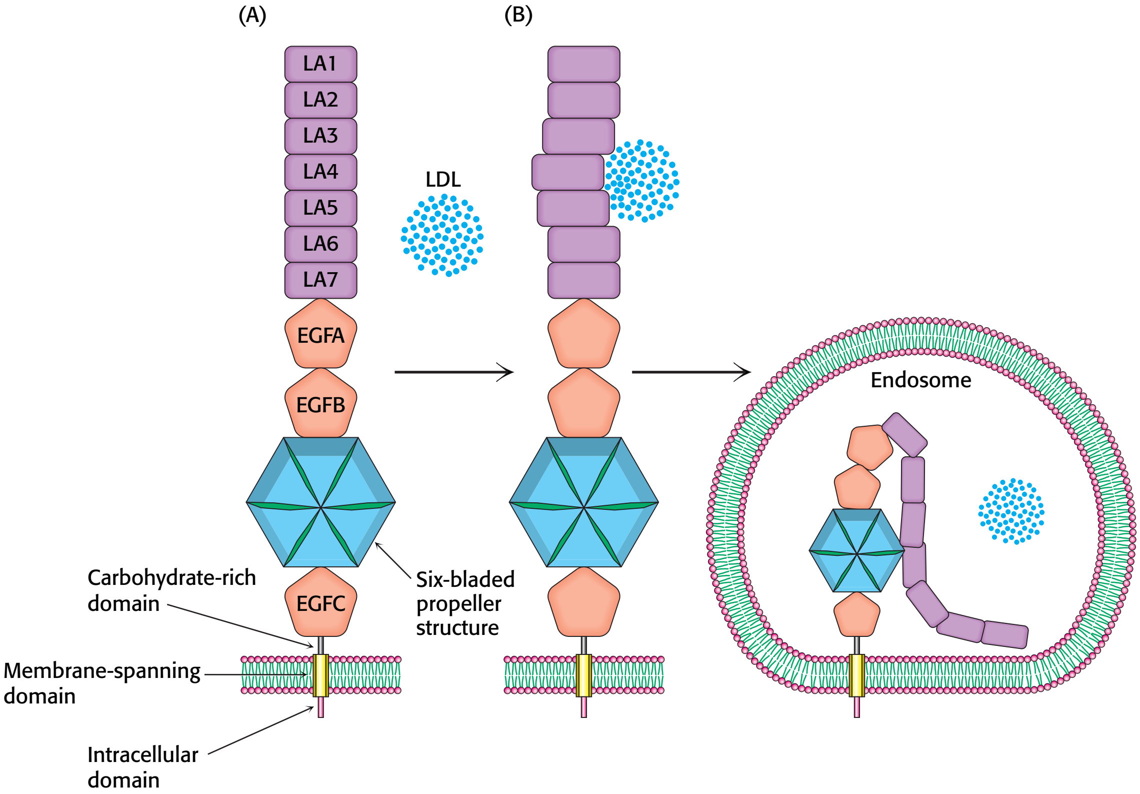


FIGURE 26.22 LDL receptor releases LDL in the endosome. (A) A schematic representation of the domain structure of the LDL receptor. (B) In the endosome, the receptor converts from the open structure into the closed structure, resulting in the release of the LDL into the endosome. [Information from I. D. Campbell, *Biochem. Soc. Trans.* 31 (pt. 6p):1107–1114, 2003, Fig. 1A.]

A second type of domain is homologous to one found in the epidermal growth factor (EGF). This domain is repeated three times, and in between the second and third repeat is a propeller structure consisting of six bladelike domains. This part of the receptor is crucial in releasing LDL in the endosome. The fourth domain, which is very rich in serine and threonine residues, contains *O*-linked sugars. These oligosaccharides function as struts to keep the receptor extended from the membrane so that the LDL-binding domain is accessible to LDL. The fifth type of domain consists of 22 hydrophobic residues that span the plasma membrane. The sixth and final domain consists of 50 residues and emerges on the cytoplasmic side of the membrane, where it controls the interaction of the receptor with coated pits and participates in endocytosis.

How does the LDL receptor relinquish its cargo on entering the endosome? The receptor exists in two interconvertible states: an extended or open state, capable of binding LDL, and a closed state that results in release of the LDL in the endosome. The receptor maintains the open state while in the plasma membrane, on binding LDL, and throughout its journey to the endosome. The conversion from the open state into the closed state takes place on exposure to the acidic environment of the endosome (Figure 26.22B). Three contiguous modules LA7, EGFA, and EGFB rigidly position the propeller module to facilitate displacement of the LDL as the closed state is formed. Under neutral pH, aspartate residues of the propeller blades form hydrogen bonds that hold each blade to the rest of the propeller structure. Exposure to the low-pH environment of the endosome causes the propeller-like structures to interact with the LDL-binding domain. This interaction displaces the LDL, which is then digested by the lysosome. The receptor is often returned to the plasma membrane to again bind LDL. The importance of this process is highlighted by the fact that more than half of the point mutations that result in familial hypercholesterolemia are due to disruptions in the interconversion between the open state and the closed state. These mutations result in a failure to release the LDL cargo and loss of the receptor by degradation.

Cycling of the LDL receptor is regulated

PCSK9 (*p*roprotein *c*onvertase *s*ubtilisin/*k*exin type) is a protease, secreted by the liver, which plays a crucial role in the regulation of cycling of the LDL receptor. Despite the fact that PCSK9 is a protease, enzymatic activity of the protein is not required for cycling regulation. PCSK9 in the blood binds to the EGFA domain of the receptor (Figure 26.22). PCSK9 locks the receptor in the open conformation even in the acidic conditions of the endosome. Failure to adopt the closed conformation prevents the receptor from returning to the plasma membrane, and it is degraded in the lysosome along with its cargo.

Individuals having a mutation that reduces the amount of PCSK9 in the blood have greatly reduced levels of LDL in the blood and display an almost 90% reduction in the rate of cardiovascular disease. Presumably, reduced levels of PCSK9 allow more receptor cycling and more efficient removal of LDL from the blood. Much research in now being directed at inhibiting PCSK9-mediated receptor degradation in individuals with high cholesterol levels. Phase 3 trials (Section 36.4) are underway, testing the efficacy of a monoclonal antibody to PCSK9. Monoclonal antibody treatments are expensive, so the search continues for a chemical inhibitor of the protein.

HDL appears to protect against atherosclerosis

Although the events that result in atherosclerosis take place rapidly in familial hypercholesterolemia, a similar sequence of events takes place in people who develop atherosclerosis over decades. In particular, the formation of foam cells and plaques are especially hazardous occurrences. HDL and its role in returning cholesterol to the liver appear to be important in mitigating these life-threatening circumstances.

HDL has a number of antiatherogenic properties, including the inhibition of LDL oxidation, but the best-characterized property is the removal of cholesterol from cells, especially macrophages. Earlier, we learned that HDL retrieves cholesterol from other tissues in the body to return the cholesterol to the liver for excretion as bile or in the feces. This transport, called *reverse cholesterol transport,* is especially important in regard to macrophages. Indeed, when the transport fails, macrophages become foam cells and facilitate the formation of plaques. Macrophages that collect cholesterol from LDL normally transport the cholesterol to HDL particles. The more HDL, the more readily this transport takes place and the less likely that the macrophages will develop into foam cells. Presumably, this robust reverse cholesterol transport accounts for the observation that higher HDL levels confer protection against atherosclerosis.

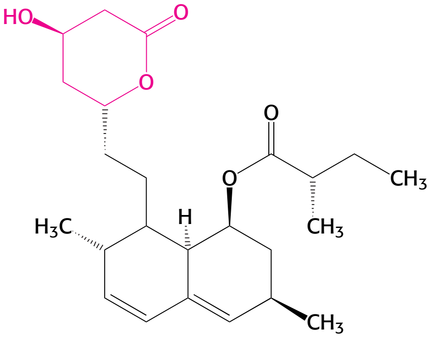
The importance of reverse cholesterol transport is illustrated by the occurrence of mutations that inactivate a cholesterol-transport protein in endothelial cells and macrophages, ABCA1 (*A*TP-*b*inding *c*assette transporter, subfamily *A*1) (Figure 13.7). Loss of activity of cholesterol-transport protein ABCA1 results in a very rare condition called *Tangier disease*, which is characterized by HDL deficiency, accumulation of cholesterol in macrophages, and premature atherosclerosis. Under normal conditions, the apoprotein component of HDL, apoA-I, binds to ABCA1 to facilitate LDL transport. Moreover, the interaction between apoA-I and ABCA1 initiates a signal transduction pathway in the endothelial cells that inhibits the inflammatory response.

Until recently, high levels of HDL-bound cholesterol ("good cholesterol") relative to LDL-bound cholesterol ("bad cholesterol") were believed to protect against cardiovascular disease. This belief was based on epidemiological studies. However, a number of recent clinical trials revealed that increased levels of HDL-bound cholesterol had no protective effects at all.

These studies do not discount the protective effects of HDL alone, but they illustrate the danger of equating free HDL and cholesterol-bound HDL.

The clinical management of cholesterol levels can be understood at a biochemical level

Homozygous familial hypercholesterolemia can be treated only by a liver transplant. A more generally applicable therapy is available for heterozygotes and others with high levels of cholesterol. *The goal is to reduce the amount of cholesterol in the blood by stimulating the synthesis of more than the customary number of LDL receptors.* We have already observed that the production of LDL receptors is controlled by the cell's need for cholesterol. The therapeutic strategy is to deprive the cell of ready sources of cholesterol. When cholesterol is required, the amount of mRNA for the LDL receptor rises and more receptors are found on the cell surface. This state can be induced by a two-pronged approach. First, the reabsorption of bile salts from the intestine is inhibited. Bile salts are cholesterol derivatives that promote the absorption of dietary cholesterol and dietary fats. Second, de novo synthesis of cholesterol is blocked.



Lovastatin

FIGURE 26.23 Lovastatin, a competitive inhibitor of HMG-CoA reductase. The part of the structure that resembles the 3-hydroxy-3-methylglutaryl moiety is shown in red.

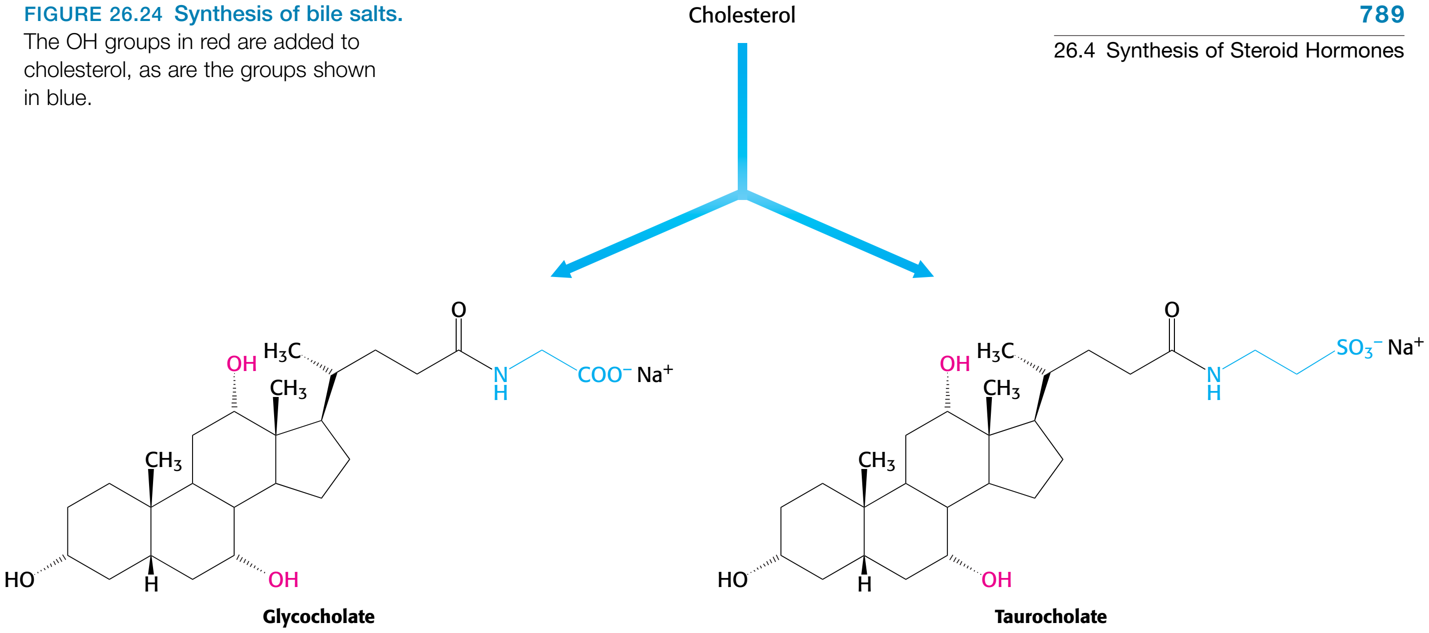
The reabsorption of bile is impeded by oral administration of positively charged polymers, such as cholestyramine, that bind negatively charged bile salts and are not themselves absorbed. Cholesterol synthesis can be effectively blocked by a class of compounds called *statins*. A well-known example of such a compound is lovastatin, which is also called mevacor (Figure 26.23). These compounds are potent competitive inhibitors ($K_i = 1$ nM) of HMG-CoA reductase, the essential control point in the biosynthetic pathway. Plasma cholesterol levels decrease by 50% in many patients given both lovastatin and inhibitors of bile-salt reabsorption. Lovastatin and other inhibitors of HMG-CoA reductase are widely used to lower the plasma-cholesterol level in people who have atherosclerosis, which is the leading cause of death in industrialized societies. Preliminary studies suggest that reducing levels of PCSK9 (p. 787) and HMG-CoA reductase activity may be an especially effective means of reducing cholesterol levels. The development of statins as effective drugs is further described in Chapter 36.

26.4 Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones

Although cholesterol is well known in its own right as a contributor to the development of heart disease, metabolites of cholesterol—the steroid hormones—also appear in the news frequently. Indeed, steroid-hormone abuse seems to be as prominent in the sports pages as any athlete is. In addition to steroid hormones, cholesterol is a precursor for two other important molecules: bile salts and vitamin D. We begin with a look at the bile salts, molecules crucial for the uptake of lipids in the diet.

Bile salts. *Bile salts* are polar derivatives of cholesterol. These compounds are highly effective *detergents* because they contain both polar and nonpolar regions. Bile salts are synthesized in the liver, stored and concentrated in the gallbladder, and then released into the small intestine. Bile salts, the major constituent of bile, *solubilize dietary lipids*. Solubilization increases the effective surface area of lipids with two consequences: (1) more surface area is exposed to the digestive action of lipases and (2) lipids are more readily absorbed by the intestine. Bile salts are also the major breakdown products of cholesterol. The bile salts glycocholate, the primary bile salt, and taurocholate are shown in Figure 26.24.

FIGURE 26.24 Synthesis of bile salts. The OH groups in red are added to cholesterol, as are the groups shown in blue.



In addition to bile salts, bile is composed of cholesterol, phospholipids, and the breakdown products of heme, bilirubin, and biliverdin (Section 24.4). If too much cholesterol is present in the bile, it will precipitate to form gallbladder stones (cholelithiasis). These stones may block bile secretion and inflame the gallbladder, a condition called cholecystitis. Symptoms include pain in the upper right abdomen, especially after a fatty meal, and nausea. If need be, the gall bladder is removed, and bile flows from the liver through the bile duct directly into the intestine.

Steroid hormones. Cholesterol is the precursor of the five major classes of *steroid hormones:* progestagens, glucocorticoids, mineralocorticoids, androgens, and estrogens (Figure 26.25). These hormones are powerful signal molecules that regulate a host of organismal functions. *Progesterone, a progestogen,* prepares the lining of the uterus for the implantation of an ovum. Progesterone is also essential for the maintenance of pregnancy by preventing premature uterine contractions. *Androgens* (such as *testosterone*) are responsible for the development of male secondary sex characteristics, whereas *estrogens* (such as *estradiol*) are required for the development of female secondary sex characteristics. Estrogens, along with progesterone, also participate in the ovarian cycle. *Glucocorticoids* (such as *cortisol*) promote gluconeogenesis and glycogen synthesis, enhance the degradation of fat and protein, and inhibit the inflammatory response. They enable animals to respond to stress; indeed, the absence of glucocorticoids can be fatal. *Mineralocorticoids* (primarily *aldosterone*) act on the distal tubules of the kidney to increase the reabsorption of Na^+ and the excretion of K^+ and H^+, which leads to an increase in blood volume and blood pressure. The major sites of synthesis of these classes of hormones are the corpus luteum, for progestogens; the testes, for androgens; the ovaries, for estrogens; and the adrenal cortex, for glucocorticoids and mineralocorticoids.

Steroid hormones bind to and activate receptor molecules that serve as transcription factors to regulate gene expression (Section 32.2). These small similar molecules

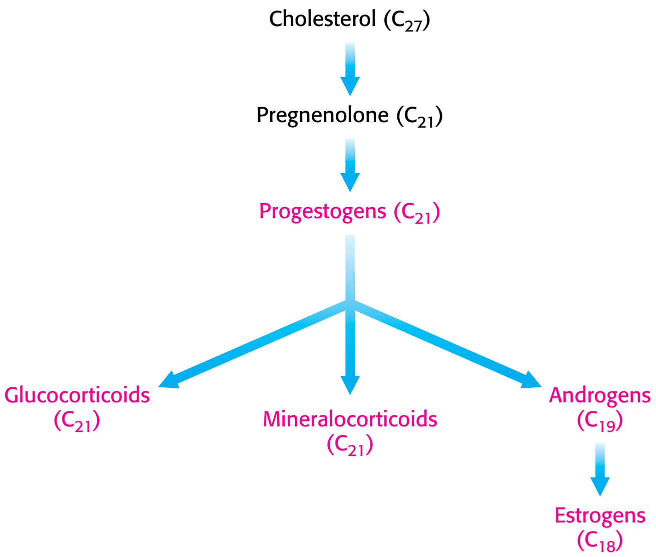


FIGURE 26.25 Biosynthetic relations of classes of steroid hormones and cholesterol.

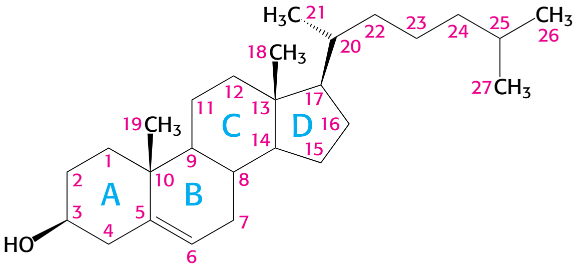
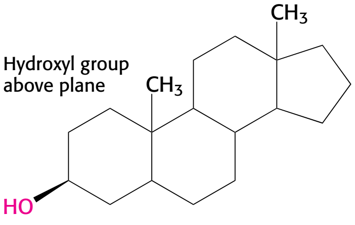


FIGURE 26.26 Cholesterol carbon numbering. The numbering scheme for the carbon atoms in cholesterol and other steroids.

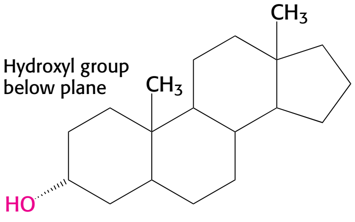
are able to have greatly differing effects because the slight structural differences among them allow interactions with specific receptor molecules.

Letters identify the steroid rings and numbers identify the carbon atoms

Carbon atoms in steroids are numbered, as shown for cholesterol in Figure 26.26. The rings in steroids are denoted by the letters A, B, C, and D. Cholesterol contains two angular methyl groups: the C-19 methyl group is attached to C-10, and the C-18 methyl group is attached to C-13. The C-18 and C-19 methyl groups of cholesterol lie *above* the plane containing the four rings. A substituent that is above the plane is termed *β oriented,* whereas a substituent that is below the plane is *α oriented.*

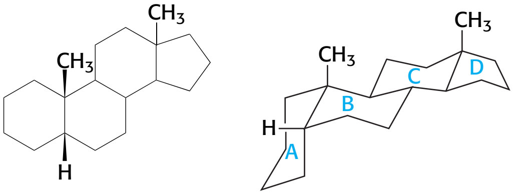


3β-Hydroxy

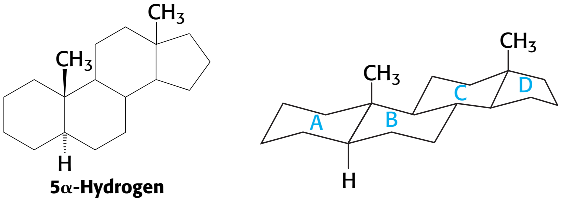


3α-Hydroxy

If a hydrogen atom is attached to C-5, it can be either α or β oriented. The A and B steroid rings are fused in a *trans* conformation if the C-5 hydrogen is *α oriented,* and *cis* if it is *β oriented.* The absence of a Greek letter for the C-5 hydrogen atom on the steroid nucleus implies a trans fusion. The C-5 hydrogen atom is α oriented in all steroid hormones that contain a hydrogen atom in that position. In contrast, bile salts have a β-oriented hydrogen atom at C-5. Thus, *a cis fusion is characteristic of the bile salts, whereas a trans fusion is characteristic of all steroid hormones that possess a hydrogen atom at C-5.* A trans fusion yields a nearly planar structure, whereas a cis fusion gives a buckled structure.



5β-Hydrogen (cis fusion)



5α-Hydrogen (trans fusion)

Steroids are hydroxylated by cytochrome P450 monooxygenases that use NADPH and O_2

The addition of OH groups plays an important role in the synthesis of cholesterol from squalene and in the conversion of cholesterol into steroid hormones and bile salts. All these hydroxylations require NADPH and O_2. The oxygen atom of the incorporated hydroxyl group comes from O_2 rather than from H_2O. Whereas one oxygen atom of the O_2 molecule goes into the substrate, the other is reduced to water. The enzymes catalyzing these reactions are called *monooxygenases* (or *mixed-function oxygenases*). Recall that a monooxygenase also participates in the hydroxylation of aromatic amino acids (Section 23.5).

$$\mathrm{RH} + \mathrm{O_2} + \mathrm{NADPH} + \mathrm{H^+} \longrightarrow \mathrm{ROH} + \mathrm{H_2O} + \mathrm{NADP^+}$$

Hydroxylation requires the activation of oxygen. In the synthesis of steroid hormones and bile salts, activation is accomplished by members of the *cytochrome P450* family, a family of cytochromes that absorb light maximally at

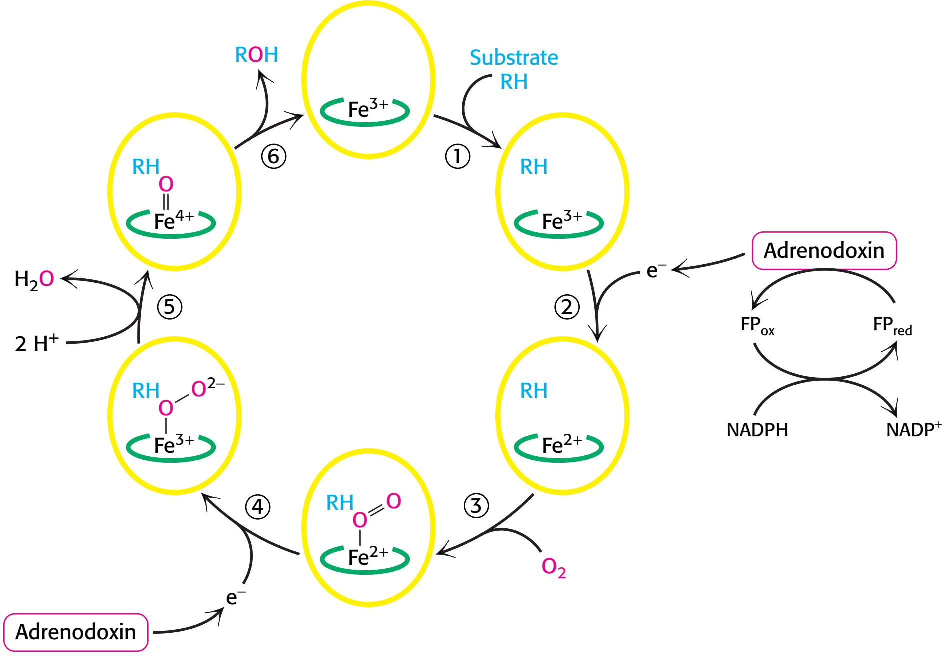


FIGURE 26.27 Cytochrome P450 mechanism. 1. Substrate binds to the enzyme. 2. Adrenodoxin donates an electron, reducing the heme iron. 3. Oxygen binds to Fe^{2+}. 4. Adrenodoxin donates a second electron. 5. The bond between the oxygen atoms is cleaved, a molecule of water is released, and an $Fe^{4+}{=}O$ intermediate is formed. 6. The $Fe^{4+}{=}O$ intermediate forms the hydroxylated product and returns the iron to the Fe^{3+} state, ready for another reaction cycle.

450 nm when complexed in vitro with exogenous carbon monoxide. These membrane-anchored proteins (~50 kDa) contain a heme prosthetic group. Oxygen is activated through its binding to the iron atom in the heme group.

Because the hydroxylation reactions promoted by P450 enzymes are oxidation reactions, it is at first glance surprising that they also consume the reductant NADPH. NADPH transfers its high-potential electrons to a flavoprotein, which transfers them, one at a time, to *adrenodoxin,* a nonheme iron protein. Adrenodoxin transfers one electron to reduce the ferric (Fe^{3+}) form of P450 to the ferrous (Fe^{2+}) form (Figure 26.27).

Without the addition of this electron, P450 will not bind oxygen. Recall that only the ferrous form (Fe^{2+}) of myoglobin and hemoglobin binds oxygen (Section 7.1). The binding of O_2 to the heme is followed by the acceptance of a second electron from adrenodoxin. The acceptance of this second electron leads to cleavage of the O—O bond. One of the oxygen atoms is then protonated and released as water. The remaining oxygen atom forms a highly reactive ferryl (Fe^{4+}) Fe═O intermediate. This intermediate abstracts a hydrogen atom from the substrate RH to form R•. This transient free radical captures the OH group from the iron atom to form ROH, the hydroxylated product, returning the iron atom to the ferric state.

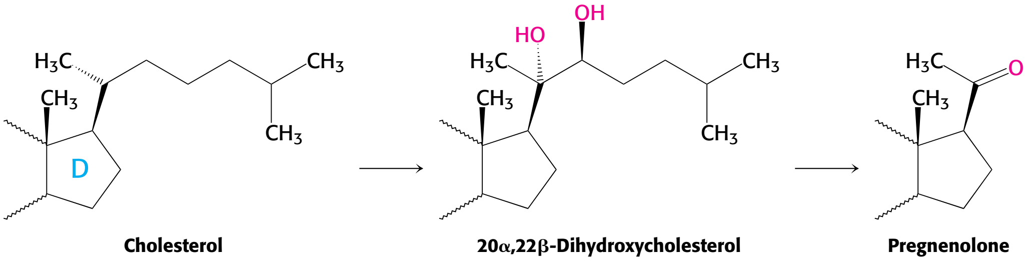
The cytochrome P450 system is widespread and performs a protective function

The cytochrome P450 system, which in mammals is located primarily in the smooth endoplasmic reticulum of the liver and small intestine, is also important in the *detoxification of foreign substances* (xenobiotic compounds). For example, the hydroxylation of phenobarbital, a barbiturate, *increases its solubility* and *facilitates its excretion.* Likewise, polycyclic aromatic hydrocarbons that are ingested by drinking contaminated water are hydroxylated by P450, providing sites for conjugation with highly polar units (e.g., glucuronate or sulfate) that markedly increase the solubility of the modified aromatic molecule. One of the most relevant functions of the cytochrome P450 system to human beings is its role in metabolizing drugs such as caffeine and ibuprofen (Chapter 36). Some members of the cytochrome P450 system also metabolize ethanol (Section 27.6). The duration of action of many medications depends on their rate of inactivation by the P450 system. Despite its general protective role in the removal of foreign chemicals, the action of the P450 system is not always beneficial. *Some of the most powerful carcinogens are generated from harmless compounds by the P450 system in vivo* in the process of *metabolic activation* (Figure 28.34). In plants, the cytochrome P450 system plays a role in the synthesis of toxic compounds as well as the pigments of flowers.

Earlier (Section 9.1), we examined the use of protease inhibitors to treat HIV infection. These inhibitors, in conjunction with inhibitors of other HIV enzymes, have drastically reduced deaths due to AIDS. The effectiveness of these drugs is sometimes compromised because P450 enzymes, such as cytochrome P450-3A4, inactivate them. The drug ritonavir was developed in the search for more effective inhibitors. Interestingly, ritonavir, although a protease inhibitor, is also a potent inhibitor of cytochrome P450-3A4. This serendipitous "side-effect" allows the clinician to administer low doses of ritonavir in combination with other protease inhibitors, effectively increasing the concentrations of these drugs while allowing a reduction in dosage and dosing frequency.

Pregnenolone, a precursor of many other steroids, is formed from cholesterol by cleavage of its side chain

Steroid hormones contain 21 or fewer carbon atoms, whereas cholesterol contains 27. Thus, the first stage in the synthesis of steroid hormones is the removal of a six-carbon unit from the side chain of cholesterol to form *pregnenolone*. The side chain of cholesterol is hydroxylated at C-20 and then at C-22, and the bond between these carbon atoms is subsequently cleaved by *desmolase*. Three molecules of NADPH and three molecules of O_2 are consumed in this six-electron oxidation.



Progesterone and corticosteroids are synthesized from pregnenolone

Progesterone is synthesized from pregnenolone in two steps. The 3-hydroxyl group of pregnenolone is oxidized to a 3-keto group, and the Δ^5 double bond is isomerized to a Δ^4 double bond (Figure 26.28). *Cortisol,* the major glucocorticoid, is synthesized from progesterone by hydroxylations at C-11, C-17, and C-21; C-17 must be hydroxylated before C-21 is hydroxylated, whereas C-11 can be hydroxylated at any stage. The enzymes catalyzing these hydroxylations are highly specific. The initial step in the synthesis of *aldosterone,* the major mineralocorticoid, is the hydroxylation of progesterone at C-21. The resulting deoxycorticosterone is hydroxylated at C-11. The oxidation of the C-18 angular methyl group to an aldehyde then yields aldosterone.

Androgens and estrogens are synthesized from pregnenolone

Androgens and estrogens also are synthesized from pregnenolone through the intermediate progesterone. Androgens contain 19 carbon atoms. The synthesis of androgens starts with the hydroxylation of progesterone at

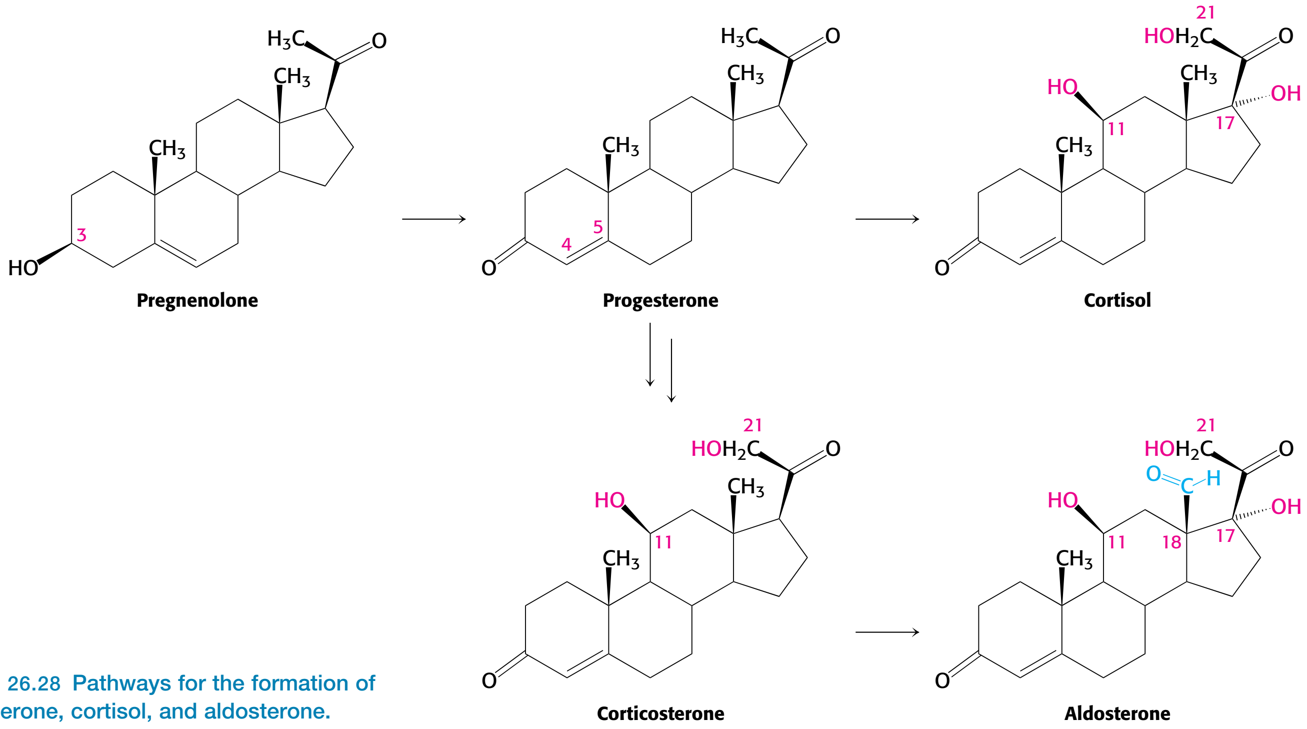
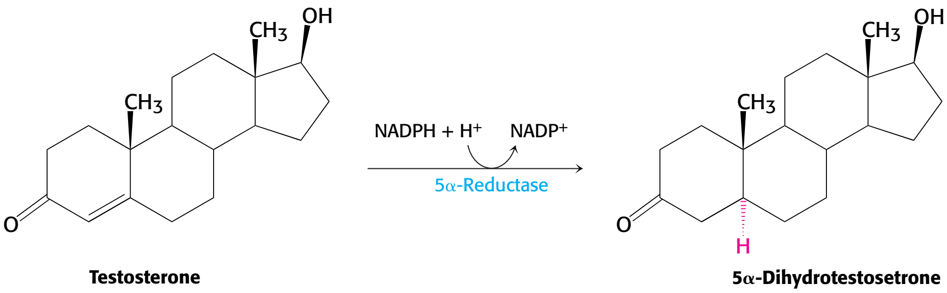


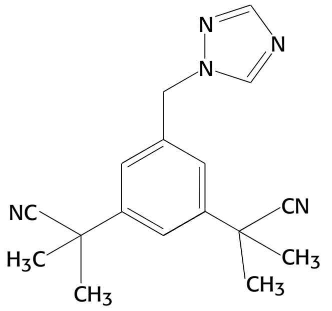
FIGURE 26.28 Pathways for the formation of progesterone, cortisol, and aldosterone.

C-17 (Figure 26.29). The side chain consisting of C-20 and C-21 is then cleaved to yield *androstenedione,* an androgen. *Testosterone,* another androgen, is formed by the reduction of the 17-keto group of androstenedione. Testosterone, through its actions in the brain, is paramount in the development of male sexual behavior. It is also important for the maintenance of the testes and the development of muscle mass. Owing to the latter activity, testosterone is referred to as an *anabolic steroid.* Testosterone is reduced by *5α-reductase* to yield *dihydrotestosterone* (DHT), a powerful embryonic androgen that instigates the development and differentiation of the male phenotype (Problems 21 and 24).

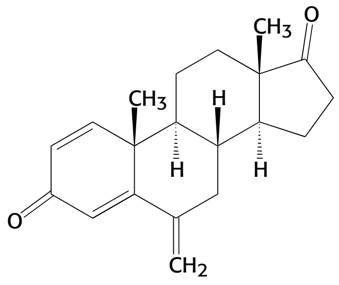


Estrogens are synthesized from androgens by the loss of the C-19 angular methyl group and the formation of an aromatic A ring. *Estrone,* an estrogen, is derived from androstenedione, whereas *estradiol,* the biologically most potent estrogen, is formed from testosterone. Estradiol can also be formed from estrone. The formation of the aromatic A ring is catalyzed by the P450 enzyme *aromatase.*

Because breast and ovarian cancers frequently depend on estrogens for growth, aromatase inhibitors are often used as a treatment for these cancers. Anastrozole is a competitive inhibitor of the enzyme, whereas exemestane is a suicide inhibitor that covalently modifies and inactivates the enzyme.



Anastrozole



Exemestane

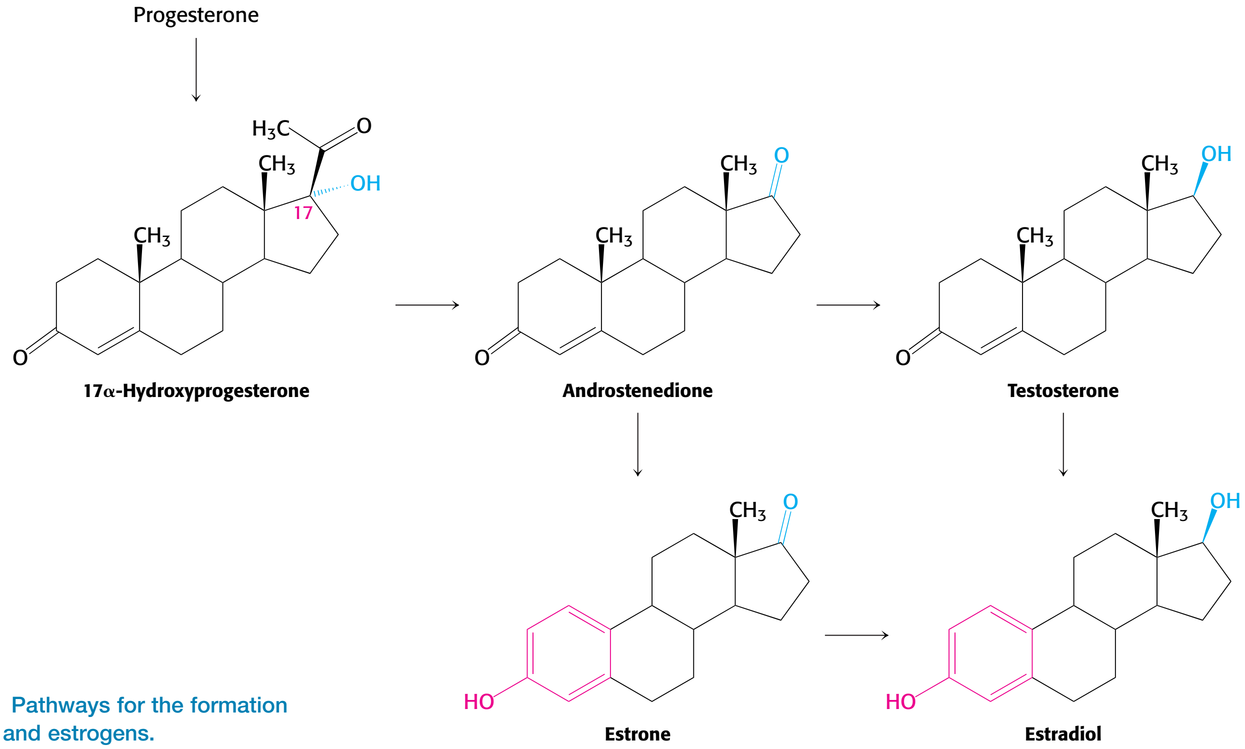

Progesterone
H3C
O
CH3
OH
17
CH3
O
17α-Hydroxyprogesterone
CH3
O
CH3
O
Androstenedione
CH3
OH
CH3
O
Testosterone
CH3
O
HO
Estrone
CH3
OH
HO
Estradiol

FIGURE 26.29 Pathways for the formation of androgens and estrogens.

Vitamin D is derived from cholesterol by the ring-splitting activity of light

Cholesterol is also the precursor of vitamin D, which plays an essential role in the control of calcium and phosphorus metabolism. *7-Dehydrocholesterol (provitamin D_3) is photolyzed by the ultraviolet light of sunlight to previtamin D_3, which spontaneously isomerizes to vitamin D_3* (Figure 26.30). Vitamin D_3

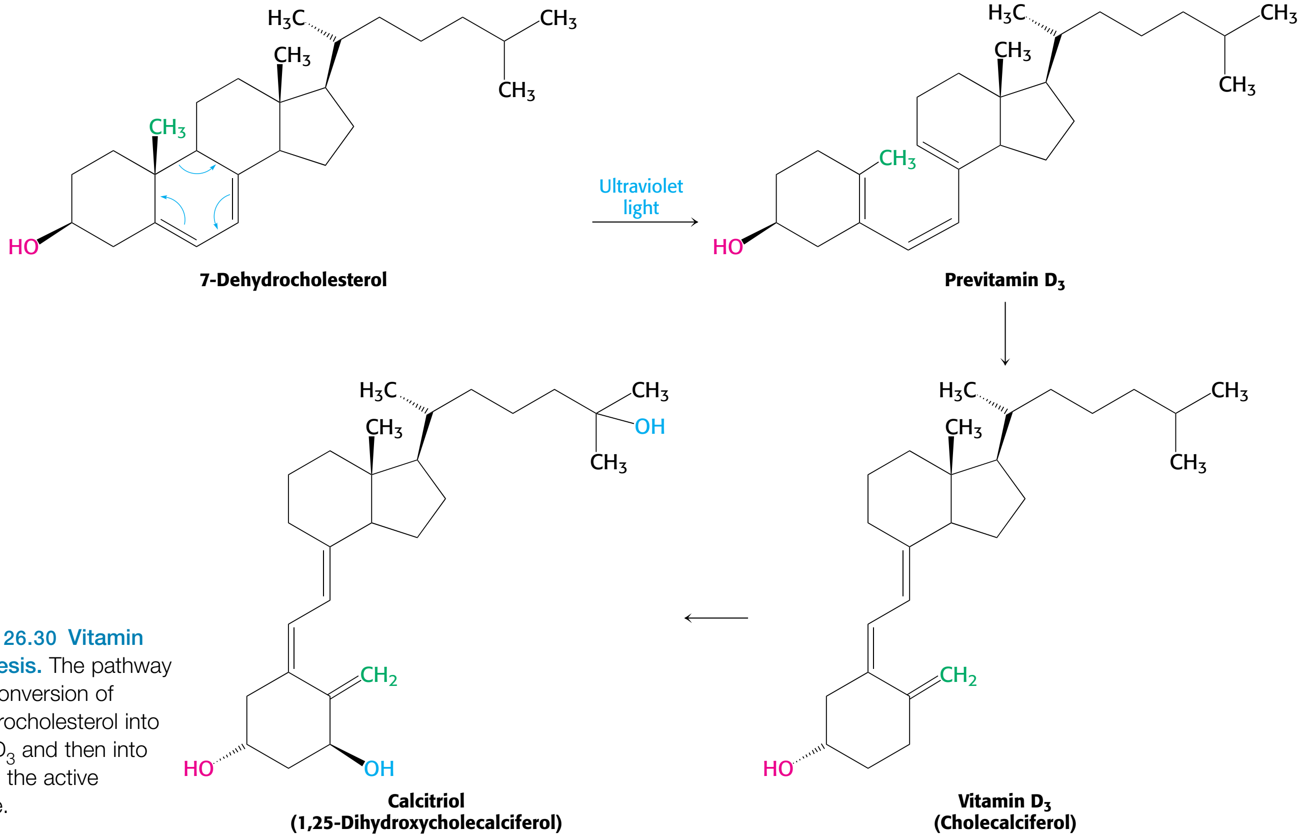

H3C
CH3
CH3
CH3
CH3
HO
7-Dehydrocholesterol
Ultraviolet light
H3C
CH3
CH3
CH3
CH3
HO
Previtamin D3
H3C
CH3
CH3
CH3
CH2
HO
Vitamin D3
(Cholecalciferol)
H3C
CH3
CH3
OH
CH3
CH2
HO
OH
Calcitriol
(1,25-Dihydroxycholecalciferol)

FIGURE 26.30 Vitamin D synthesis. The pathway for the conversion of 7-dehydrocholesterol into vitamin D_3 and then into calcitriol, the active hormone.

(cholecalciferol) is converted into *calcitriol* (1,25-dihydroxycholecalciferol), the active hormone, by hydroxylation reactions in the liver and kidneys. Although not a steroid, vitamin D acts in an analogous fashion. It binds to a receptor, structurally similar to the steroid receptors, to form a complex that functions as a transcription factor, regulating gene expression.

Vitamin D deficiency in childhood produces *rickets,* a disease characterized by inadequate calcification of cartilage and bone. Rickets was so common in seventeenth-century England that it was called the "children's disease of the English." The 7-dehydrocholesterol in the skin of these children was not photolyzed to previtamin D_3, because there was little sunlight for many months of the year. Furthermore, their diets provided little vitamin D, because most naturally occurring foods have a low content of this vitamin. Fish-liver oils are a notable exception. Cod-liver oil, abhorred by generations of children because of its unpleasant taste, was used in the past as a rich source of vitamin D. Today, the most reliable dietary sources of vitamin D are fortified foods. Milk, for example, is fortified to a level of 400 international units per quart (10 μg per quart). The recommended daily intake of vitamin D is 200 international units until age 50, after which it increases with age. In adults, vitamin D deficiency leads to softening and weakening of bones, a condition called *osteomalacia.* The occurrence of osteomalacia in Muslim women who are clothed so that only their eyes are exposed to sunlight is a striking reminder that vitamin D is needed by adults as well as by children.

Research over the past few years indicates that vitamin D may play a much larger biochemical role than simply the regulation of bone metabolism. Muscle seems to be a target of vitamin D action. In muscle, vitamin D appears to affect a number of biochemical processes with the net effect being enhanced muscle performance. Studies also suggest that vitamin D prevents cardiovascular disease, reduces the incidence of a variety of cancers, and protects against autoimmune diseases including diabetes. Moreover, vitamin D deficiency appears to be more common than previously thought. In the United States, 75% of Blacks and many Hispanics and Asians have insufficient blood levels of vitamin D. This recent research on vitamin D shows again the dynamic nature of biochemical investigations. Vitamin D, a chemical whose biochemical role was believed to be well established, now offers new frontiers of biomedical research.

SUMMARY

26.1 Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols

Phosphatidate is formed by successive acylations of glycerol 3-phosphate by acyl CoA. Hydrolysis of its phosphoryl group followed by acylation yields a triacylglycerol. CDP-diacylglycerol, the activated intermediate in the de novo synthesis of several phospholipids, is formed from phosphatidate and CTP. The activated phosphatidyl unit is then transferred to the hydroxyl group of a polar alcohol, such as inositol, to form a phospholipid such as phosphatidylinositol. In mammals, phosphatidylethanolamine is formed by CDP-ethanolamine and diacylglycerol. Phosphatidylethanolamine is methylated by *S*-adenosylmethionine to form phosphatidylcholine. In mammals, this phosphoglyceride can also be synthesized by a pathway that utilizes dietary choline. CDP-choline is the activated intermediate in this route.

Sphingolipids are synthesized from ceramide, which is formed by the acylation of sphingosine. Gangliosides are sphingolipids that contain an oligosaccharide unit having at least one residue of *N*-acetylneuraminate or a related sialic acid. They are synthesized by the step-by-step addition of activated sugars, such as UDP-glucose, to ceramide.

26.2 Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages

Cholesterol is a steroid component of animal membranes and a precursor of steroid hormones. The committed step in its synthesis is the formation of mevalonate from 3-hydroxy-3-methylglutaryl CoA (derived from acetyl CoA and acetoacetyl CoA). Mevalonate is converted into isopentenyl pyrophosphate (C_5), which condenses with its isomer, dimethylallyl pyrophosphate (C_5), to form geranyl pyrophosphate (C_{10}). The addition of a second molecule of isopentenyl pyrophosphate yields farnesyl pyrophosphate (C_{15}), which condenses with itself to form squalene (C_{30}). This intermediate cyclizes to lanosterol (C_{30}), which is modified to yield cholesterol (C_{27}).

26.3 The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels

In the liver, cholesterol synthesis is regulated by changes in the amount and activity of 3-hydroxy-3-methylglutaryl CoA reductase. Transcription of the gene, translation of the mRNA, and degradation of the enzyme are stringently controlled. In addition, the activity of the reductase is regulated by phosphorylation.

Triacylglycerols exported by the intestine are carried by chylomicrons and then hydrolyzed by lipases lining the capillaries of target tissues. Cholesterol and other lipids in excess of those needed by the liver are exported in the form of very low density lipoprotein. After delivering its content of triacylglycerols to adipose tissue and other peripheral tissue, VLDL is converted into intermediate-density lipoprotein and then into low-density lipoprotein. IDL and LDL carry cholesteryl esters, primarily cholesteryl linoleate. Liver and peripheral tissue cells take up LDL by receptor-mediated endocytosis. The LDL receptor, a protein spanning the plasma membrane of the target cell, binds LDL and mediates its entry into the cell. Absence of the LDL receptor in the homozygous form of familial hypercholesterolemia leads to a markedly elevated plasma level of LDL cholesterol and the deposition of cholesterol on blood-vessel walls, which in turn may result in childhood heart attacks. High-density lipoproteins transport cholesterol from the peripheral tissues to the liver.

26.4 Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones

In addition to bile salts, which facilitate the digestion of lipids, five major classes of steroid hormones are derived from cholesterol: progestogens, glucocorticoids, mineralocorticoids, androgens, and estrogens. Hydroxylations by P450 monooxygenases that use NADPH and O_2 play an important role in the synthesis of steroid hormones and bile salts from cholesterol. P450 enzymes, a large superfamily, also participate in the detoxification of drugs and other foreign substances.

Pregnenolone (C_{21}) is an essential intermediate in the synthesis of steroids. This steroid is formed by scission of the side chain of cholesterol. Progesterone (C_{21}), synthesized from pregnenolone, is the precursor of cortisol and aldosterone. Hydroxylation of progesterone and cleavage of its side chain yields androstenedione, an androgen (C_{19}).

Estrogens (C_{18}) are synthesized from androgens by the loss of an angular methyl group and the formation of an aromatic A ring. Vitamin D, which is important in the control of calcium and phosphorus metabolism, is formed from a derivative of cholesterol by the action of light.

KEY TERMS

phosphatidate (p. 768)
triacylglycerol (p. 769)
phospholipid (p. 769)
cytidine diphosphodiacylglycerol (CDP-diacylglycerol) (p. 769)
sphingolipid (p. 772)
ceramide (p. 772)
cerebroside (p. 772)
ganglioside (p. 772)
cholesterol (p. 776)
mevalonate (p. 776)
3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase) (p. 776)
3-isopentenyl pyrophosphate (p. 777)
sterol regulatory element binding protein (SREBP) (p. 779)
lipoprotein particles (p. 782)
low-density lipoprotein (LDL) (p. 782)
high-density lipoprotein (HDL) (p. 782)
receptor-mediated endocytosis (p. 784)
reverse cholesterol transport (p. 787)
bile salt (p. 788)
steroid hormone (p. 789)
cytochrome P450 monooxygenase (p. 790)

PROBLEMS

1. *Different roles.* Describe the roles of glycerol 3-phosphate, phosphatidate, and diacylglycerol in triacylglycerol and phospholipid synthesis.

2. *Needed supplies.* How is the glycerol 3-phosphate required for phosphatidate synthesis generated?

3. *Making fat.* Write a balanced equation for the synthesis of a triacylglycerol, starting from glycerol and fatty acids.

4. *Making a phospholipid.* Write a balanced equation for the synthesis of phosphatidylethanolamine by the de novo pathway, starting from ethanolamine, glycerol, and fatty acids.

5. *ATP needs.* How many high-phosphoryl-transfer-potential molecules are required to synthesize phosphatidylethanolamine from ethanolamine and diacylglycerol? Assume that the ethanolamine is the activated component.

6. *Identifying differences.* Differentiate among sphingomyelin, a cerebroside, and a ganglioside.

7. *Let's count the ways.* There may be 50 ways to leave your lover, but, in principle, there are only three ways to make a glycerol-based phospholipid. Describe the three pathways.

8. *Activated donors.* What is the activated reactant in each of the following biosyntheses?

(a) Phosphatidylinositol from inositol

(b) Phosphatidylethanolamine from ethanolamine

(c) Ceramide from sphingosine

(d) Sphingomyelin from ceramide

(e) Cerebroside from ceramide

(f) Ganglioside G_{M1} from ganglioside G_{M2}

(g) Farnesyl pyrophosphate from geranyl pyrophosphate

9. *No DAG, no TAG.* What would be the effect of a mutation that decreased the activity of phosphatidic acid phosphatase?

10. *Like Wilbur and Orville.* Match each term with its description.

(a) Phosphatidate ____
(b) Triacylglycerol ____
(c) Phospholipid ____
(d) Sphingolipid ____
(e) Cerebroside ____
(f) Ganglioside ____
(g) Cholesterol ____
(h) Mevalonate ____
(i) Lipoprotein particle ____
(j) Steroid hormone ____

1. Glycerol-based membrane lipid
2. Product of the committed step in cholesterol synthesis
3. Ceramide with either glucose or galactose attached
4. Storage form of fatty acids
5. Squalene is a precursor to this molecule
6. Transports cholesterol and lipids
7. Derived from cholesterol
8. Precursor to both phospholipids and triacylglycerols
9. Formed from ceramide by the attachment of phosphocholine
10. Ceramide with multiple carbohydrates attached

11. *The Law of Three Stages.* What are the three stages required for the synthesis of cholesterol?

12. *Many regulations to follow.* Outline the mechanisms of the regulation of cholesterol biosynthesis.

13. *Telltale labels.* What is the distribution of isotopic labeling in cholesterol synthesized from each of the following precursors?

(a) Mevalonate labeled with ^{14}C in its carboxyl carbon atom

(b) Malonyl CoA labeled with ^{14}C in its carboxyl carbon atom

14. *Too much, too soon*. What is familial hypercholesterolemia and what are its causes?

15. *Familial hypercholesterolemia.* Several classes of LDL-receptor mutations have been identified as causes of this disease. Suppose that you have been given cells from patients with different mutations, an antibody specific for the LDL receptor that can be seen with an electron microscope, and access to an electron microscope. What differences in antibody distribution might you expect to find in the cells from different patients?

16. *Breakfast conversation*. You and a friend are eating breakfast together. While eating, your friend is reading the back of her cereal box and comes across the following statement: "Cholesterol plays beneficial roles in your body, making cells, hormones, and tissues." Knowing that you are taking biochemistry, she asks if the statement makes sense. What do you reply?

17. *A good thing*. What are statins? What is their pharmacological function?

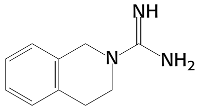
18. *Too much of a good thing*. Would the development of a "super statin" that inhibited all HMG CoA reductase activity be a useful drug? Explain.

19. *RNA editing.* A shortened version (apo B-48) of apolipoprotein B is formed by the intestine, whereas the full-length protein (apo B-100) is synthesized by the liver. A glutamine codon (CAA) is changed into a stop codon. Propose a simple mechanism for this change.

20. *A means of entry*. Describe the process of receptor-mediated endocytosis by using LDL as an example.

21. *Inspiration for drug design.* Some actions of androgens are mediated by dihydrotestosterone, which is formed by the reduction of testosterone. This finishing touch is catalyzed by an NADPH-dependent 5α-reductase (p. 793). Chromosomal XY males with a genetic deficiency of this reductase are born with a male internal urogenital tract but predominantly female external genitalia. These people are usually reared as girls. At puberty, they masculinize because the testosterone level rises. The testes of these reductase-deficient men are normal, whereas their prostate glands remain small. How might this information be used to design a drug to treat *benign prostatic hypertrophy,* a common consequence of the normal aging process in men? A majority of men older than age 55 have some degree of prostatic enlargement, which often leads to urinary obstruction.

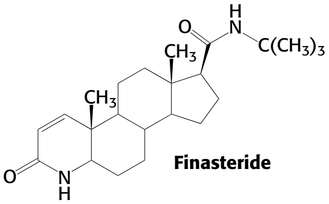
22. *Drug idiosyncrasies.* Debrisoquine, a β-adrenergic blocking agent, has been used to treat hypertension. The optimal dose varies greatly (20–400 mg daily) in a population of patients. The urine of most patients taking the drug contains a high level of 4-hydroxydebrisoquine. However, those most sensitive to the drug (about 8% of the group studied) excrete debrisoquine and very little of the 4-hydroxy derivative. Propose a molecular basis for this drug idiosyncrasy. Why should caution be exercised in giving other drugs to patients who are very sensitive to debrisoquine?



Debrisoquine

23. *Removal of odorants.* Many odorant molecules are highly hydrophobic and concentrate within the olfactory epithelium. They would give a persistent signal independent of their concentration in the environment if they were not rapidly modified. Propose a mechanism for converting hydrophobic odorants into water-soluble derivatives that can be rapidly eliminated.

24. *Development difficulties.* Propecia (finasteride) is a synthetic steroid that functions as a competitive and specific inhibitor of 5α-reductase (p. 793), the enzyme responsible for the synthesis of dihydrotestosterone from testosterone.



Finasteride

In addition to its use as a treatment for benign prostatic hypertrophy (Problem 21), finasteride is also widely used to retard the development of male pattern hair loss. Pregnant women are advised to avoid handling this drug. Why is it vitally important that pregnant women avoid contact with Propecia?

25. *Life-style consequences.* Human beings and the plant *Arabidopsis* evolved from the same distant ancestor possessing a small number of cytochrome P450 genes. Human beings have approximately 50 such genes, whereas *Arabidopsis* has more than 250 of them. Propose a role for the large number of P450 isozymes in plants.

26. *Personalized medicine.* The cytochrome P450 system metabolizes many medicinally useful drugs. Although all human beings have the same number of P450 genes, individual polymorphisms exist that alter the specificity and efficiency of the proteins encoded by the genes. How could knowledge of individual polymorphisms be useful clinically?

27. *Honeybee crisis*. In 2006, there was a sudden, unexplained die-off of honeybee colonies throughout the United States. The die-off was economically significant because one-third of the human diet comes from insect-pollinated plants, and honeybees are responsible for 80% of the pollination. In October of 2006, the sequence of the honeybee

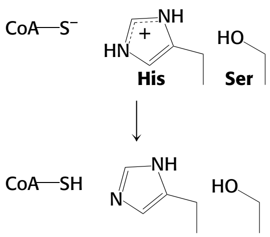
genome was reported. Interestingly, the genome was found to contain far fewer cytochrome P450 genes than do the genomes of other insects. Suggest how the die-off and the paucity of P450 genes may be related.

28. *Missing enzyme*. Congenital adrenal hyperplasia is a life-threatening condition that results from a deficiency in the P450 enzyme steroid 21-hydroxylase. This enzyme catalyzes the first step in the conversion of progesterone into cortisol and aldosterone (Figures 26.28 and 26.29). A characteristic of congenital adrenal hyperplasia is an increase in sex hormone production. Explain why this is the case.

29. *Let the sun shine in*. At a biochemical level, vitamin D functions like a steroid hormone (Chapter 31). Therefore, it is sometimes referred to as an honorary steroid. Why is vitamin D not an actual steroid?

Mechanism Problems

30. *An interfering phosphate*. In the course of the overall reaction catalyzed by HMG-CoA reductase, a histidine residue protonates a coenzyme A thiolate, CoA– S^-, generated in an earlier step.



A nearby serine residue can be phosphorylated by AMP-dependent kinase, which results in a loss of activity. Propose an explanation for why phosphorylation of the serine residue inhibits enzyme activity.

31. *Demethylation*. Methyl amines are often demethylated by cytochrome P450 enzymes. Propose a mechanism for the formation of methylamine from dimethylamine catalyzed by cytochrome P450. What is the other product?

Chapter Integration Problems

32. *Similarities*. Compare the role of CTP in phosphoglyceride synthesis with the role of UTP in glycogen synthesis.

33. *Hold on tight or you might be thrown to the cytoplasm*. Many proteins are modified by the covalent attachment of a farnesyl (C_{15}) or a geranylgeranyl (C_{20}) unit to the carboxyl-terminal cysteine residue of the protein. Suggest why this modification might occur.

34. *Fork in the road*. 3-Hydroxy-3-methylglutaryl CoA is on the pathway for cholesterol biosynthesis. It is also a component of another pathway. Name the pathway. What determines which pathway 3-hydroxy-3-methylglutaryl CoA follows?

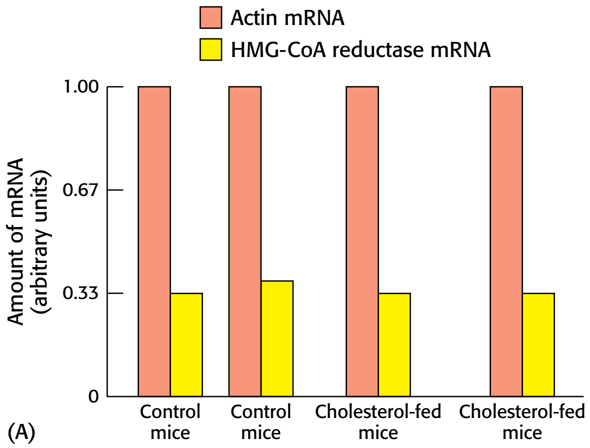
35. *Requires a club membership*. How is methionine metabolism related to the synthesis of phosphatidylcholine?

36. *Drug resistance*. Dichlorodiphenyltrichloroethane (DDT) is a potent insecticide rarely used today because of its effects on other forms of life. In insects, DDT disrupts sodium channel function, leading to eventual death. Mosquitos have developed resistance to DDT and other insecticides that function in a similar fashion. Suggest two means by which DDT resistance might develop.

37. *ATP requirements*. Explain how cholesterol synthesis depends on the activity of ATP-citrate lyase.

Data Interpretation and Chapter Integration Problem

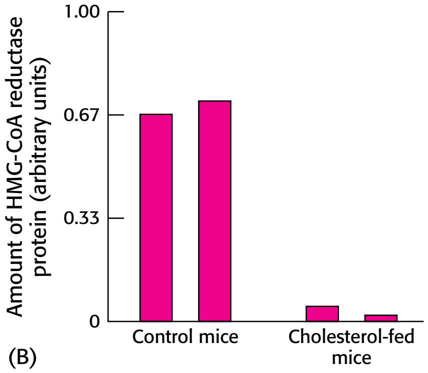
38. *Cholesterol feeding*. Mice were divided into four groups, two of which were fed a normal diet and two of which were fed a cholesterol-rich diet. HMG-CoA reductase mRNA and protein from liver were then isolated and quantified. Graph A shows the results of the mRNA isolation.



(a) What is the effect of cholesterol feeding on the amount of HMG-CoA reductase mRNA?

(b) What is the purpose of also isolating the mRNA for the protein actin, which is not under the control of the sterol regulatory element?

HMG-CoA reductase protein was isolated by precipitation with a monoclonal antibody to HMG-CoA reductase. The



amount of HMG-CoA protein in each group is shown in graph B.

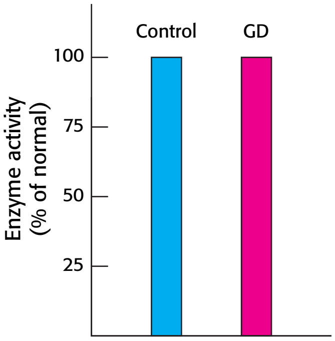
(c) What is the effect of the cholesterol diet on the amount of HMG-CoA reductase protein?

(d) Why is this result surprising in light of the results in graph A?

(e) Suggest possible explanations for the results in graph B.

39. *Gaucher disease*. Gaucher disease, the most common lysosomal storage disease in humans, is caused by mutations in the gene encoding glucocerebrosidase (GCase), the lysosomal enzyme that degrades glucocerebrosides. The disease has a variety of symptoms depending on the severity of the disease and the person affected. Identical twins can display very different levels of severity. Symptoms include bone pain, enlarged liver, excessive fatigue, and mental retardation.

In order to better understand the disease, researchers undertook a series of experiments to characterize the nature of the defect in the enzyme. Cells were obtained from a person without the disease (control) as well as from a patient with the disease (GD). The cells were grown in culture, the enzyme was isolated, and enzyme activity of 10 μg of each sample was measured (Figure A).

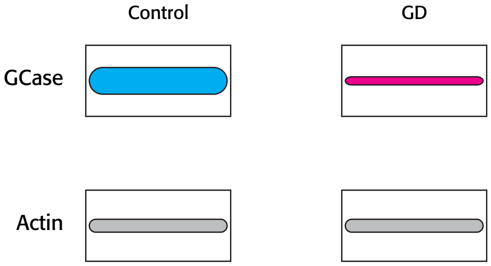


(A)

[Data from J. Lu, Proc. Natl. Acad. Sci. U. S. A. 107:21665–21670, 2012.]

(a) What do the results tell about the catalytic activity of the enzyme from the GD cells? Why were these results surprising to the researchers?

Again, cells were grown in culture and cell extracts were made from the control and GD samples. Care was taken to insure that the same number of cells were used in both extractions. A western blot with antibodies to GCase was performed on the cell extracts, with the results shown in Figure B.



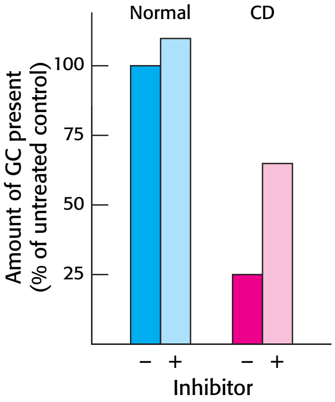
(B)

[Data from J. Lu, Proc. Natl. Acad. Sci. U. S. A. 107:21665–21670, 2012.]

(b) Provide two possible explanations for the result shown in Figure B. Why was it crucial to insure that the extract was made from the same number of cells? What was the purpose of including the western blot of actin?

(c). The researchers then measured the amount of mRNA for the enzyme in both the control and GD cells. The amount of mRNA was identical in the two samples. With this information, reinterpret the results shown in Figure B.

Next, cells from the control and GD individuals were grown in the presence and absence of a potent proteasome inhibitor. The amount of GCase was determined using western blots. The results are shown in Figure C.



(C)

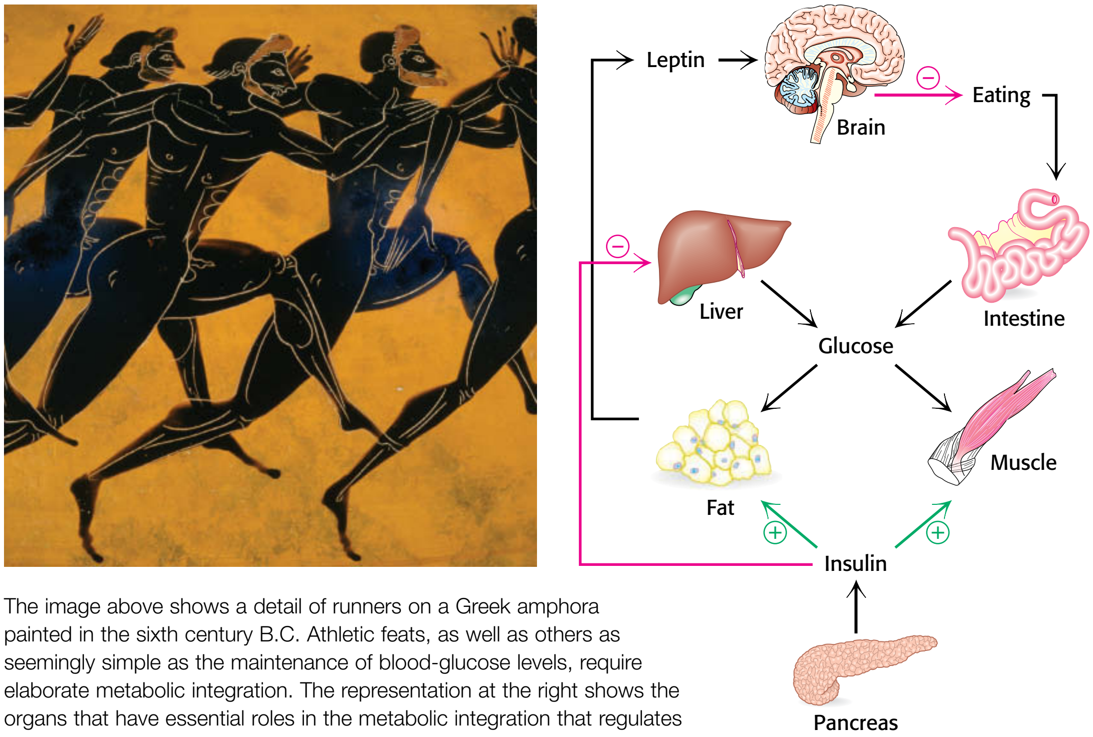
[Data from J. Lu, Proc. Natl. Acad. Sci. U. S. A. 107:21665–21670, 2012.]

(d) Suggest the nature of the defect in the GD enzyme. What is the significance of the increase in activity of the enzyme from the normal cell observed in the presence of the inhibitor?

CHAPTER

27

The Integration of Metabolism



The image above shows a detail of runners on a Greek amphora painted in the sixth century B.C. Athletic feats, as well as others as seemingly simple as the maintenance of blood-glucose levels, require elaborate metabolic integration. The representation at the right shows the organs that have essential roles in the metabolic integration that regulates blood-glucose levels during exercise and at rest. Insulin and leptin (secreted by adipocytes) are two of the hormones that modulate the metabolic pathways of organs throughout the body such that adequate energy is available to meet the demands of living. [(Left) Copyright © The Metropolitan Museum of Art/Art Resource, NY.]

We have been examining the biochemistry of metabolism one pathway at a time. We have seen how useful energy is extracted from fuels and used to power biosynthetic reactions and signal-transduction pathways. In Chapters 28 through 30, we will extend our study of biosynthetic reactions to the synthesis of nucleic acids and proteins. Before we do that, however, in this chapter we will take a step back to examine large-scale biochemical interactions that constitute the physiology of the organism. In keeping with a central theme of life—energy manipulations—we will look at the regulation of energy at the organismal level, which boils down to an apparently simple but actually quite complex question: At the biochemical level, how does an organism know when to eat and when to refrain from eating? The ability to maintain adequate but not excessive energy stores is called *caloric homeostasis* or *energy homeostasis*.

Next, we will examine a significant perturbation of caloric homeostasis—obesity—and how this physiological condition affects insulin action, frequently resulting in diabetes. We then turn our attention to a biochemical consideration of one of the most beneficial activities that humans can

OUTLINE

27.1 Caloric Homeostasis Is a Means of Regulating Body Weight
27.2 The Brain Plays a Key Role in Caloric Homeostasis
27.3 Diabetes Is a Common Metabolic Disease Often Resulting from Obesity
27.4 Exercise Beneficially Alters the Biochemistry of Cells
27.5 Food Intake and Starvation Induce Metabolic Changes
27.6 Ethanol Alters Energy Metabolism in the Liver

engage in—exercise—and see how exercise mitigates the effects of diabetes as well as how different forms of exercise use different sources of fuel.

At the opposite end of the physiological spectrum from obesity and overnutrition are fasting and starvation, and we will examine the biochemical responses to these challenges. The chapter ends with a consideration of another biochemical energy perturbation—excess alcohol consumption.

We have already encountered instances of organismal energy regulation when we considered the actions of insulin and glucagon. Recall that insulin, secreted by the β cells of the pancreas, causes glucose to be removed from the blood and stimulates the synthesis of glycogen and lipids. Glucagon, secreted by the α cells of the pancreas, has effects opposite those of insulin. Glucagon increases the level of blood glucose by stimulating glycogen breakdown and gluconeogenesis. This chapter introduces two other hormones that play key roles in caloric homeostasis. Leptin and adiponectin, secreted by the adipose tissue, work in concert with insulin to regulate caloric homeostasis.

27.1 Caloric Homeostasis Is a Means of Regulating Body Weight

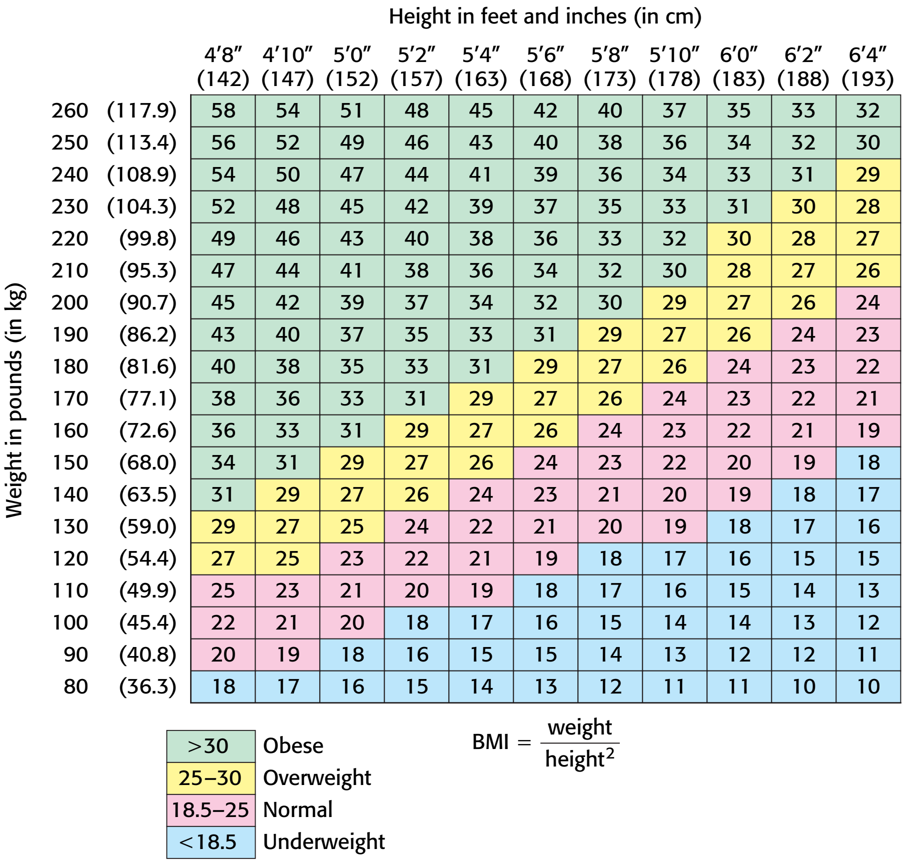
By now in our study of biochemistry, we are well aware of the fact that carbohydrates and lipids are sources of energy. We consume these energy sources as foods, convert the energy into ATP, and use the ATP to power our lives. Like all energy transformations, our energy consumption and expenditure are governed by the laws of thermodynamics. Recall that the First Law of Thermodynamics states that energy can neither be created nor destroyed. Translated into the practical terms of our diets,

$$\text{Energy consumed} = \text{energy expended} + \text{energy stored}$$

This simple equation has severe physiological and health implications: according to the First Law of Thermodynamics, if we consume more energy than we expend, we will become overweight or obese. Obesity is generally defined as a body mass index (BMI) of more than 30 kg m^{-2}, whereas overweight is defined as a BMI of more than 25 kg m^{-2} (Figure 27.1). Recall that excess fat is stored in adipocytes as triacylglycerols. The number of adipocytes becomes fixed in adults, and so obesity results in engorged adipocytes. Indeed, the cells may increase as much as 1000-fold in size.

We are all aware that many of us, especially in the developed world, are becoming obese or have already attained that state. In the United States, obesity has become an epidemic, with nearly 30% of adults classified as such. Obesity is identified as a risk factor in a host of pathological conditions including diabetes mellitus, hypertension, and cardiovascular disease (Table 27.1). The cause of obesity is quite simple in most cases: more food is consumed than is needed, and the excess calories are stored as fat. We will consider the biochemical basis of pathologies caused by obesity later in the chapter.

Before we undertake a biochemical analysis of the results of overconsumption, let us consider why the obesity epidemic is occurring in the first place. There are several possible, overlapping explanations. The first is a commonly held view that our bodies are programmed to rapidly store excess calories in times of plenty, an evolutionary adaptation from times past when humans were not assured of having ample food, as many of us are today. Consequently, we store calories as if a fast might begin tomorrow, but no such fast arrives. A second possible explanation is that we no longer face the risks of predation. Evidence indicates that predation was a common cause of death for our ancestors. An obese individual would more likely have been culled from a group of our ancestors than would a more nimble, lean individual. As the risk of predation declined, leanness became less beneficial.


Height in feet and inches (in cm)
Weight in pounds (in kg)

Weight in pounds (in kg)	4′8″ (142)	4′10″ (147)	5′0″ (152)	5′2″ (157)	5′4″ (163)	5′6″ (168)	5′8″ (173)	5′10″ (178)	6′0″ (183)	6′2″ (188)	6′4″ (193)
260 (117.9)	58	54	51	48	45	42	40	37	35	33	32
250 (113.4)	56	52	49	46	43	40	38	36	34	32	30
240 (108.9)	54	50	47	44	41	39	36	34	33	31	29
230 (104.3)	52	48	45	42	39	37	35	33	31	30	28
220 (99.8)	49	46	43	40	38	36	33	32	30	28	27
210 (95.3)	47	44	41	38	36	34	32	30	28	27	26
200 (90.7)	45	42	39	37	34	32	30	29	27	26	24
190 (86.2)	43	40	37	35	33	31	29	27	26	24	23
180 (81.6)	40	38	35	33	31	29	27	26	24	23	22
170 (77.1)	38	36	33	31	29	27	26	24	23	22	21
160 (72.6)	36	33	31	29	27	26	24	23	22	21	19
150 (68.0)	34	31	29	27	26	24	23	22	20	19	18
140 (63.5)	31	29	27	26	24	23	21	20	19	18	17
130 (59.0)	29	27	25	24	22	21	20	19	18	17	16
120 (54.4)	27	25	23	22	21	19	18	17	16	15	15
110 (49.9)	25	23	21	20	19	18	17	16	15	14	13
100 (45.4)	22	21	20	18	17	16	15	14	14	13	12
90 (40.8)	20	19	18	16	15	15	14	13	12	12	11
80 (36.3)	18	17	16	15	14	13	12	11	11	10	10

>30	Obese
25–30	Overweight
18.5–25	Normal
<18.5	Underweight

$$\text{BMI} = \frac{\text{weight}}{\text{height}^2}$$

FIGURE 27.1 Body mass index (BMI). The BMI value for an individual person is a reliable indicator of obesity for most people. [Data from the Centers for Disease Control.]

A third possibility, one currently receiving much attention, is that calorie-dense highly palatable foods—foods rich in sugar and fats—that are readily accessible to most people in the developed world, act as drugs, triggering the same reward pathways in the brain that are triggered by drugs such as cocaine. These reward pathways can be strong enough to override appetite-suppressing signals. Fourth, a growing body of research suggests that our intestinal microbiome—the bacteria that inhibit our guts—plays a significant role in how we process our food. For instance, germ-free mice do not become obese, even when given unfettered access to a high-fat diet. However, when such mice are exposed to the intestinal flora of obese mice, the mice now become obese even on a normal chow diet. Moreover, the intestinal microbiome of obese mice triggers an inflammatory response that may blunt the effect of signal molecules that normally regulate the desire to eat as well as how the mice process the calories they consume. Many of these results have been extended to humans. Finally, it is evident that individuals respond differently to obesity-inducing environmental conditions and that

TABLE 27.1 Health consequences of obesity or being overweight

Coronary heart disease
Type 2 diabetes
Cancers (endometrial, breast, colon and others)
Hypertension (high blood pressure)
Dyslipidemia (disruption of lipid metabolism, e.g., high cholesterol and triacylglycerols)
Stroke
Liver and gallbladder disease
Sleep apnea and respiratory problems
Osteoarthritis (degeneration of cartilage and underlying bone at a joint)
Gynecological problems (abnormal menses, infertility)

Information from: Centers for Disease Control and Prevention Web site (www.cdc.gov).

this difference has a large genetic component. Various studies have indicated the heritability of fat mass to be between 30% and 70%, indicating that the tendency toward obesity may be highly heritable.

Regardless of why we may have a propensity to gain weight, this propensity can be counteracted behaviorally—by eating less and exercising more. It also appears that it is easier to prevent weight gain than it is to lose weight. Clearly, caloric homeostasis is a complicated biological phenomenon. Understanding this phenomenon will engage biomedical research scientists for some time to come.

As disturbing as the obesity epidemic is, an equally intriguing, almost amazing observation is that many people are able to maintain an approximately constant weight throughout adult life, despite consuming tons of food over a lifetime (Problems 1 and 2). Although will power, exercise, and a bathroom scale often play a role in this homeostasis, some biochemical signaling must be taking place to achieve this remarkable physiological feat.

27.2 The Brain Plays a Key Role in Caloric Homeostasis

What makes this remarkable balance of energy input and output possible? As one might imagine, the answer is complicated, entailing many biochemical signals as well as a host of behavioral factors. We will focus on a few key biochemical signals, and divide our discussion into two parts: short-term signals that are active during a meal and long-term signals that report on the overall energy status of the body. These signals originate in the gastrointestinal tract, the β cells of the pancreas, and fat cells. The primary target of these signals is the brain, in particular a groups of neurons in a region of the hypothalamus called the arcuate nucleus.

Signals from the gastrointestinal tract induce feelings of satiety

Short-term signals relay feelings of satiety from the gut to various regions of the brain and thus reduce the urge to eat (Figure 27.2). The best-studied short-term signal is cholecystokinin (CCK). *Cholecystokinin* is actually a family of peptide hormones of various lengths (from 8 to 58 amino acids in length, depending on posttranslational processing) secreted into the blood by cells in the duodenum and jejunum regions of the small intestine as a

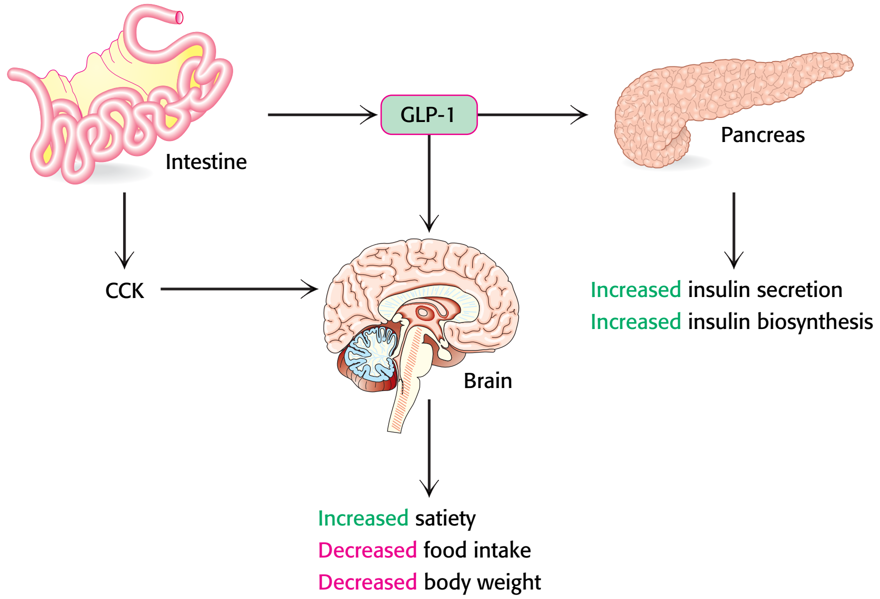


FIGURE 27.2 Satiation signals. Cholecystokinin (CCK) and glucagon-like peptide 1 (GLP-1) are signal molecules that induce feelings of satiety in the brain. CCK is secreted by specialized cells of the small intestine in response to a meal and activates satiation pathways in the brain. GLP-1, secreted by L cells in the intestine, also activates satiation pathways in the brain and potentiates insulin action in the pancreas. [Information from S. C. Wood, *Cell Metab*. 9:489–498, 2009, Fig. 1.]

postprandial satiation signal. The CCK binds to the CCK receptor, a G-protein-coupled receptor (Section 14.1) located in various peripheral neurons, that relay signals to the brain. This binding initiates a signal-transduction pathway in the brain that generates a feeling of satiety. CCK also plays an important role in digestion, stimulating the secretion of pancreatic enzymes and bile salts from the gallbladder.

Another important gut signal is *glucagon-like peptide 1* (GLP-1), a hormone of approximately 30 amino acids in length. GLP-1 is secreted by intestinal L cells, hormone-secreting cells located throughout the lining of the gastrointestinal tract. GLP-1 has a variety of effects, all apparently facilitated by binding to its receptor, another G-protein-coupled receptor. Like CCK, GLP-1 induces feelings of satiety that inhibit further eating. GLP-1 also potentiates glucose-induced insulin secretion by the β cells of the pancreas while inhibiting glucagon secretion.

Although we have examined only two short-term signals, many others are believed to exist (Table 27.2). Most of the short-term signals thus far identified are appetite suppressants. Ghrelin, a peptide that is 28 amino acids in length and secreted by the stomach, acts on regions of the hypothalamus to stimulate appetite through its receptor, a G-protein-coupled receptor. Ghrelin secretion increases before a meal and decreases afterward.

TABLE 27.2 Gastrointestinal peptides that regulate food intake

Appetite-suppressing signals
Cholecystokinin
Glucagon-like peptide 1
Glucagon-like peptide 2
Amylin
Somatostatin
Bombesin
Enterostatin
Apolipoprotein A-IV
Gastric inhibitory peptide
Appetite-enhancing peptides
Ghrelin

Information from: M. H. Stipanuk, ed., *Biochemical, Physiological, Molecular Aspects of Human Nutrition,* 2d ed. (Saunders/Elsevier, 2006), p. 627, Box 22-1.

Leptin and insulin regulate long-term control over caloric homeostasis

Two key signal molecules regulate energy homeostasis over the time scale of hours or days: *leptin,* which is secreted by the adipocytes, and *insulin,* which is secreted by the β cells of the pancreas. Leptin reports on the status of the triacylglycerol stores, whereas insulin reports on the status of glucose in the blood—in other words, of carbohydrate availability. We will consider leptin first.

Adipose tissue was formerly considered an inert depot of triacylglycerols. However, recent work has shown that adipose tissue is an active endocrine tissue, secreting signal molecules called *adipokines*, such as leptin, that regulate a host of physiological processes. Leptin is secreted by the adipocytes in direct proportion to the amount of fat present. The more fat in a body, the more leptin is secreted. Leptin binding to its receptor throughout the body increases the sensitivity of muscle and the liver to insulin, stimulates β oxidation of fatty acids, and decreases triacylglycerol synthesis.

Let us consider the effects of leptin in the brain. Leptin binds to its receptor, thereby activating a signal transduction pathway. The leptin receptor is found in various regions of the brain, but particularly in the arcuate nucleus of the hypothalamus. There, one population of neurons expresses appetite-stimulating (orexigenic) peptides, called neuropeptide Y (NPY) and agouti-related peptide (AgRP). Fasting stimulates the production of NPY and AgRP owing to the decrease in leptin levels that results from diminishing adipose tissue (Figure 27.3A). Leptin, on the other hand, inhibits the NPY/AgRP neurons, preventing the release of NPY and AgRP and thus repressing the desire to eat (Figure 27.3B).

The second population of neurons containing leptin receptors expresses a large precursor polypeptide, proopiomelanocortin (POMC). In response to leptin binding to its receptor on POMC neurons, POMC is proteolytically processed to yield a variety of signal molecules, one of which, *melanocyte-stimulating hormone* (MSH), is especially important in this context. MSH, originally discovered as a stimulator of melanocytes (cells that synthesize the pigment melanin), activates appetite-suppressing (anorexigenic) neurons and thus inhibits food consumption. Fasting inhibits MSH activity and thus stimulates eating. AgRP inhibits MSH activity by acting as an antagonist, binding to the MSH receptor but failing to activate the receptor (Figure 27.3). Thus, the net effect of leptin binding to its receptor is the initiation of a complex signal-transduction pathway that ultimately curtails food intake.

Insulin receptors are also present in the hypothalamus, although the mechanism of insulin action in the brain is less clear than that of leptin. Insulin appears to inhibit NPY/AgRP-producing neurons, thus inhibiting food consumption.

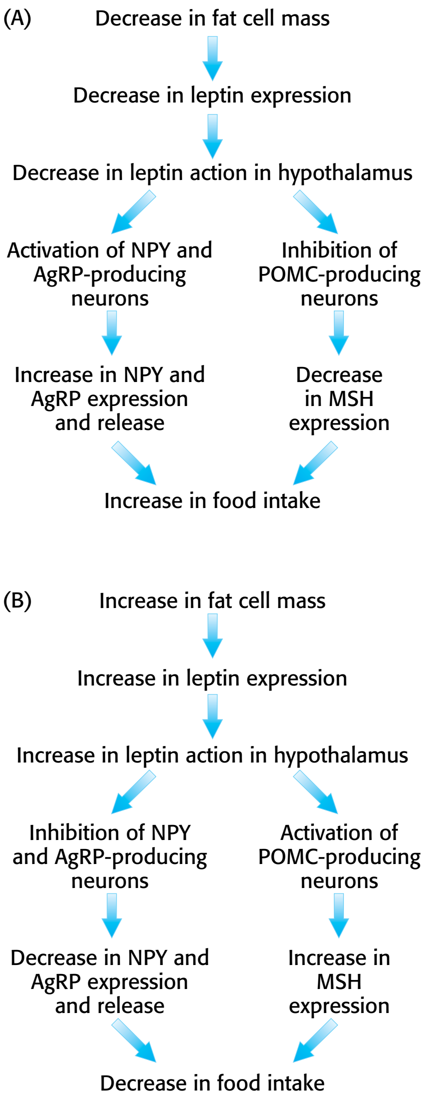

(A)
Decrease in fat cell mass
Decrease in leptin expression
Decrease in leptin action in hypothalamus
Activation of NPY and AgRP-producing neurons
Inhibition of POMC-producing neurons
Increase in NPY and AgRP expression and release
Decrease in MSH expression
Increase in food intake
(B)
Increase in fat cell mass
Increase in leptin expression
Increase in leptin action in hypothalamus
Inhibition of NPY and AgRP-producing neurons
Activation of POMC-producing neurons
Decrease in NPY and AgRP expression and release
Increase in MSH expression
Decrease in food intake

FIGURE 27.3 The effects of leptin in the brain. Leptin is an adipokine secreted by adipose tissue in direct relation to fat mass. (A) When leptin levels fall, as in fasting, appetite-enhancing neuropeptides NPY and AgRP are secreted, whereas the secretion of appetite-suppressing signals such as MSH is inhibited. (B) When fat mass increases, leptin inhibits NPY and AgRP secretion while stimulating the release of appetite-suppressing hormone MSH. [Information from M. H. Stipanuk, *Biochemical, Physiological, & Molecular Aspects of Human Nutrition,* 2d ed. (Saunders/Elsevier, 2006), Fig. 22-2.]

Leptin is one of several hormones secreted by adipose tissue

Leptin was the first adipokine discovered because of the dramatic effects of its absence. Researchers discovered a strain of mice called ob/ob mice, which lack leptin and, as a result, are extremely obese (refer to Chapter 22 opening figure on p. 643). These mice display hyperphagia (overeating), hyperlipidemia (accumulation of triacylglycerols in muscle and liver), and an insensitivity to insulin. Since the discovery of leptin, other adipokines have been detected. For instance, *adiponectin* is another signal molecule produced by the adipocytes. Secretion of adiponectin falls in direct proportion to increases in fat mass. A key function of adiponectin appears to be increasing the sensitivity of the organism to insulin. Both leptin and adiponectin exert their effects through the key regulatory enzyme, AMP-activated protein kinase (AMPK). Recall that this enzyme is active when AMP levels are elevated and ATP levels are diminished, and this activation leads to a decrease in anabolism and an increase in catabolism, most notably an increase in fatty acid oxidation (Section 22.6). In insulin-resistant obese animals such as the ob/ob mice, leptin levels increase while those of adiponectin decrease.

Adipocytes also produce two hormones, *RBP4* (originally discovered as a *r*etinol *b*inding *p*rotein) and *resistin*, that promote insulin resistance. Although it is unclear why adipocytes secrete hormones that facilitate insulin resistance, a pathological condition, we can speculate on the reason. These signal molecules may help to fine-tune the actions of leptin and adiponectin or perhaps to act as "brakes" on the action of leptin and adiponectin to prevent hypoglycemia in the fasted state. Some evidence indicates that enlarged adipocytes that result from obesity may secrete higher levels of insulin-antagonizing hormones and thus contribute to insulin resistance. Resistin has recently been implicated as a causal factor in the increased incidence of cardiovascular disease in obese individuals.

Leptin resistance may be a contributing factor to obesity

If leptin is produced in proportion to body-fat mass and leptin inhibits eating, why do people become obese? Obese people, in most cases, have both functioning leptin receptors and a high blood concentration of leptin. The failure to respond to the anorexigenic effects of leptin is called *leptin resistance*. What is the basis of leptin resistance?

As for most questions in the exciting area of energy homeostasis, the answer is not well worked out, but recent evidence suggests that a group of proteins called *suppressors of cytokine signaling* (SOCS) may take part. These proteins fine-tune some hormonal systems by inhibiting receptor action. SOCS proteins inhibit receptor signaling by a number of means. Consider, for example, the effect of SOCS proteins on the insulin receptor. Recall that insulin stimulates the autophosphorylation of tyrosine residues on the insulin receptor, which in turn phosphorylates IRS-1, initiating the insulin-signaling pathway (Figure 27.4A). SOCS proteins bind to phosphorylated tyrosine residues on receptors or other members of the signal-transduction pathway, thereby disrupting signal flow and thus altering the cell's biochemical activity (Figure 27.4B). In other cases, the binding of SOCS proteins to components

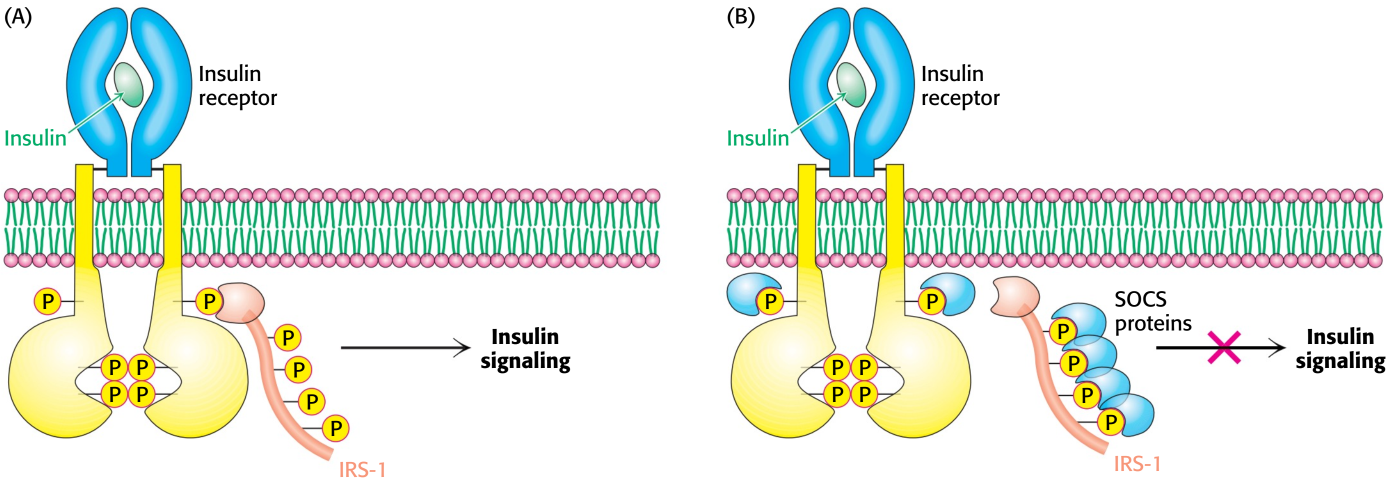


FIGURE 27.4 Suppressors of cytokine signaling (SOCS) regulate receptor function. (A) Insulin binding results in phosphorylation of the receptor and subsequent phosphorylation of IRS-1. These processes initiate the insulin-signaling pathway. (B) SOCS proteins disrupt interactions of components of the insulin-signaling pathway by binding phosphorylated proteins and thereby inhibiting the pathway. The binding of a signal component by SOCS results in proteasomal degradation in some cases. (IRS-1, insulin-receptor substrate 1; SOCS, suppressor of cytokine signaling.)

of the signal-transduction pathway may also enhance proteolytic degradation of these components by the proteasome (Section 23.2). Evidence in support of a role for SOCS in leptin resistance comes from mice that have had SOCS selectively deleted from POMC-expressing neurons. These mice display an enhanced sensitivity to leptin and are resistant to weight gain even when fed a high-fat diet. The reason why the activity of SOCS proteins increases, leading to leptin resistance, remains to be determined.

Dieting is used to combat obesity

Given the obesity epidemic that we currently face and its associated disorders, much attention has been focused on determining the most effective weight-loss diet. In general, two categories of diet try to help us control our caloric intake—low-carbohydrate diets and low-fat diets. Low-carbohydrate diets usually emphasize protein consumption. Although studies of the effects of diets on humans are immensely complex, data are beginning to accumulate suggesting that low-carbohydrate–high-protein diets *may* be the most effective for losing weight. The exact reasons are not clear, but there are two common hypotheses. First, proteins seem to induce a feeling of satiation more effectively than do fats or carbohydrates. Second, proteins require more energy to digest than do fats or carbohydrates, and the increased energy expenditure contributes to weight loss. For instance, some studies show that a diet that is 30% protein requires almost 30% more energy to digest than that required by a diet that is 10% protein. The mechanisms by which protein-rich diets enhance energy expenditure and feelings of satiation remain to be determined. Regardless of the type of diet, the adage "Eat less, exercise more" always applies.

27.3 Diabetes Is a Common Metabolic Disease Often Resulting from Obesity

Having taken an overview of the regulation of body weight, we now examine the biochemical results when regulation fails because of behavior, genetics, or a combination of the two. The most common result of such a failure is obesity, a condition in which excess energy is stored as triacylglycerols. Recall that all excess food consumption is ultimately converted into triacylglycerols. Humans maintain about a day's worth of glycogen and, after these stores have been replenished, excess carbohydrates are converted into fats and then into triacylglycerols. Amino acids are not stored at all, and so excess amino acids are ultimately converted into fats also. Thus, regardless of the type of food consumed, excess consumption results in increased fat stores.

We begin our consideration of the effects of disruptions in caloric homeostasis with *diabetes mellitus,* a complex disease characterized by grossly abnormal fuel usage: *glucose is overproduced by the liver and underutilized by other organs*. The incidence of diabetes mellitus (usually referred to simply as *diabetes*) is about 5% of the world population. *Type 1 diabetes* is caused by the autoimmune destruction of the insulin-secreting β cells in the pancreas and usually begins before age 20. Type 1 diabetes is also referred to as insulin-dependent diabetes, meaning that the affected person requires the administration of insulin to live. Most diabetics, in contrast, have a normal or even higher level of insulin in their blood, but they are unresponsive to the hormone, a characteristic called *insulin resistance*. This form of the disease, known as *type 2 diabetes,* typically arises later in life than does the insulin-dependent form. Type 2 diabetes accounts for approximately 90% of the diabetes cases throughout the world and is the most common metabolic disease in the world. In the United States, it is the leading cause of blindness, kidney failure, and amputation. *Obesity is a significant predisposing factor for the development of type 2 diabetes.*

Diabetes

Named for the excessive urination in the disease. Aretaeus, a Cappadocian physician of the second century A.D., wrote: "The epithet diabetes has been assigned to the disorder, being something like passing of water by a siphon." He perceptively characterized diabetes as "being a melting-down of the flesh and limbs into urine."

Mellitus

From Latin, meaning "sweetened with honey." Refers to the presence of sugar in the urine of patients having the disease. *Mellitus* distinguishes this disease from diabetes *insipidus,* which is caused by impaired renal reabsorption of water.

Insulin initiates a complex signal-transduction pathway in muscle

What is the biochemical basis of insulin resistance? How does insulin resistance lead to failure of the β cells of the pancreas that results in type 2 diabetes? How does obesity contribute to this progression? To answer these questions and begin to unravel the mysteries of metabolic disorders, let us examine the mechanism of action of insulin in muscle, the largest tissue target of insulin. Indeed, muscle uses about 85% of the glucose ingested during a meal.

In a normal cell, insulin binds to its receptor, which then autophosphorylates on tyrosine residues, with each subunit of the receptor phosphorylating its partner. Phosphorylation of the receptor generates binding sites for insulin-receptor substrates (IRSs), such as IRS-1 (Figure 27.5). Subsequent phosphorylation of IRS-1 by the tyrosine kinase activity of the insulin receptor engages the insulin-signaling pathway (1). Phosphorylated IRS-1 binds to phosphoinositide 3-kinase (PI3K) and activates it. PI3K catalyzes the conversion of phosphatidylinositol 4,5-bisphosphate (PIP_2) into phosphatidylinositol 3,4,5-trisphosphate

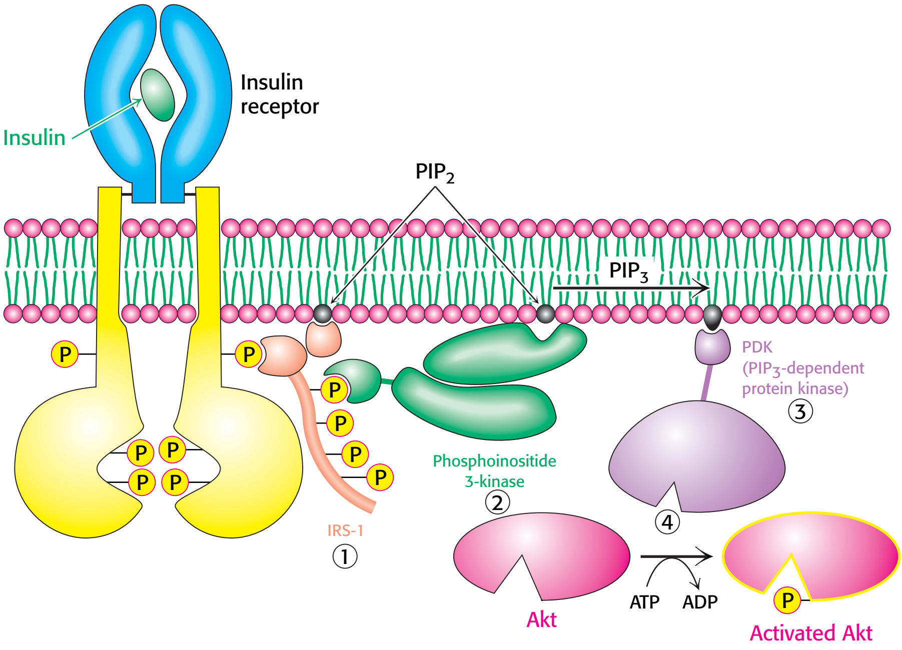
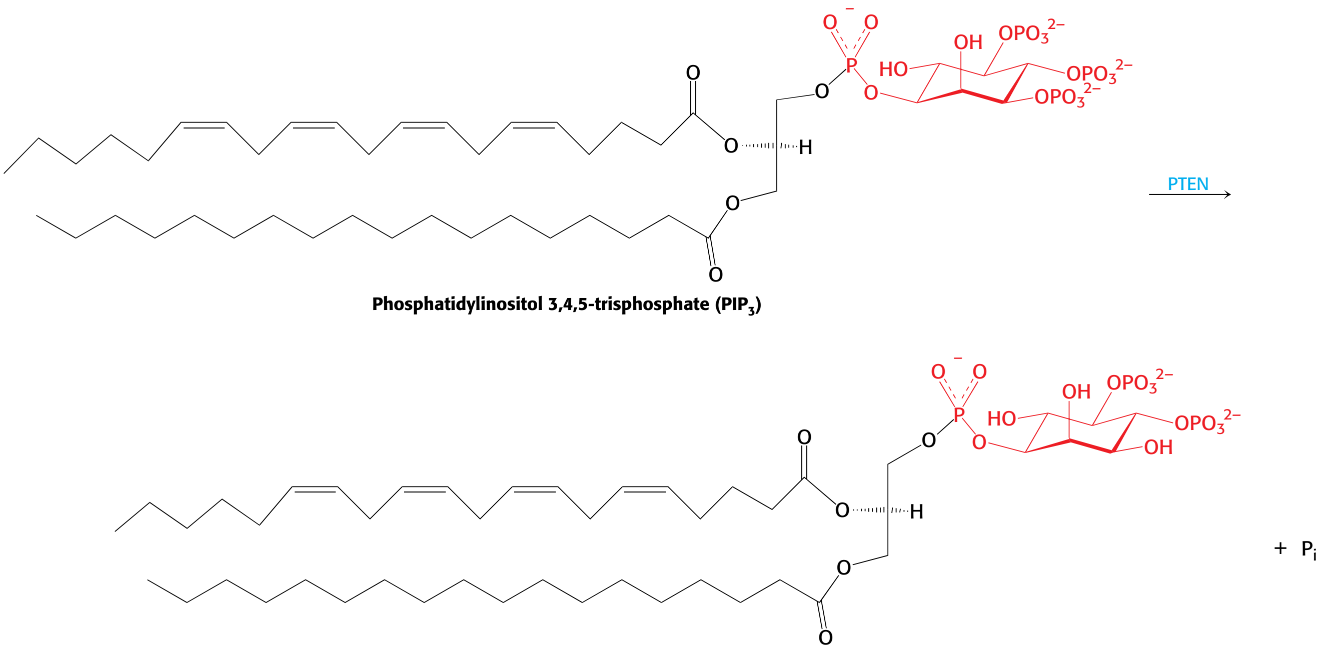


FIGURE 27.5 Insulin signaling. The binding of insulin results in the cross-phosphorylation and activation of the insulin receptor. Phosphorylated sites on the receptor act as binding sites for insulin-receptor substrates such as IRS-1. The lipid kinase phosphoinositide 3-kinase binds to phosphorylated sites on IRS-1 through its regulatory domain and then converts PIP_2 into PIP_3. Binding to PIP_3 activates PIP_3-dependent protein kinase, which phosphorylates and activates kinases such as Akt. Activated Akt can then diffuse throughout the cell and continue the signal-transduction pathway.

(PIP$_3$), a second messenger (2). PIP$_3$ activates the phosphatidylinositol-dependent protein kinase (PDK) (3), which in turn activates several other kinases, most notably Akt (4), also known as protein kinase B (PKB). Akt facilitates the translocation of the glucose transporter (GLUT4)-containing vesicles to the cell membrane, which leads to a more robust absorption of glucose from the blood. Moreover, Akt phosphorylates and inhibits glycogen synthase kinase (GSK3). Recall that GSK3 inhibits glycogen synthase (Section 21.5). Thus, insulin leads to the absorption of glucose from the blood, activation of glycogen synthase, and enhanced glycogen synthesis.

Like all signal pathways, the insulin-signaling cascade must be capable of being turned off. Three different processes contribute to the down-regulation of insulin signaling. First, phosphatases deactivate the insulin receptor and destroy a key second messenger. *Tyrosine phosphatase IB* removes phosphoryl groups from the receptor, thus inactivating it. The second messenger PIP$_3$ is inactivated by the phosphatase *PTEN* (*p*hospha-*t*ase and *ten*sin homolog), which dephosphorylates it, forming PIP$_2$, which itself has no second-messenger properties.



Phosphatidylinositol 3,4,5-trisphosphate (PIP$_3$)

Phosphatidylinositol 4,5-bisphosphate (PIP$_2$)

Second, the IRS protein can be inactivated by phosphorylation on serine residues by specific Ser/Thr kinases. These kinases are activated by overnutrition and other stress signals and may play a role in the development of insulin resistance. Finally, SOCS proteins, the regulatory proteins discussed earlier, interact with the insulin receptor and IRS-1 and apparently facilitate their proteolytic degradation by the proteasome complex.

Metabolic syndrome often precedes type 2 diabetes

With our knowledge of the key components of energy homeostasis, let us begin our investigation of the biochemical basis of insulin resistance and type 2 diabetes. Obesity is a contributing factor to the development of insulin resistance, which is an early development on the path to type 2 diabetes. Indeed, a cluster of pathologies—including insulin resistance, hyperglycemia, dyslipidemia (high blood levels of triacylglycerols, cholesterol, and low-density lipoproteins)—often develop together. This clustering, called *metabolic syndrome,* is thought to be a predecessor of type 2 diabetes.

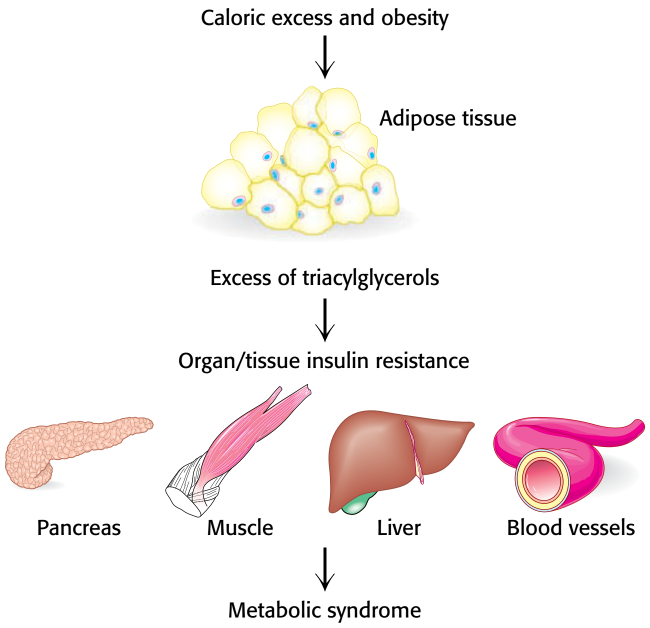

Caloric excess and obesity
Adipose tissue
Excess of triacylglycerols
Organ/tissue insulin resistance
Pancreas
Muscle
Liver
Blood vessels
Metabolic syndrome

FIGURE 27.6 The storage capacity of fat tissue can be exceeded in obesity. In caloric excess, the storage capacity of adipocytes can be exceeded with deleterious results. The excess fat accumulates in other tissues, resulting in biochemical malfunction of the tissues. When the pancreas, muscle, liver, and cells lining the blood vessels are affected, metabolic syndrome, a condition that often precedes type 2 diabetes, may result. [Information from S. Fröjdö, H. Vidal, and L. Pirola, *Biochim. Biophys. Acta* 1792:83–92, 2009, Fig. 1.]

A consequence of obesity is that the amount of triacylglycerols consumed exceeds the adipose tissue's storage capacity. As a result, other tissues begin to accumulate fat, most notably liver and muscle (Figure 27.6). For reasons to be presented later in the chapter, this accumulation results in insulin resistance and ultimately in pancreatic failure. We will focus on muscle and the β cells of the pancreas.

Excess fatty acids in muscle modify metabolism

We have seen many times the importance of fats as fuels. In regard to obesity, more fats are present than can be processed by muscle. Although the rate of β oxidation increases in response to the high concentration of fats, mitochondria are not capable of processing all of the fatty acids by β oxidation; fatty acids accumulate in the mitochondria and eventually spill over into the cytoplasm. Indeed, the inability to process all of the fatty acids results in their reincorporation into triacylglycerols and the accumulation of fat in the cytoplasm. In the cytoplasm, levels of diacylglycerol and ceramide (a component of sphingolipids) also increase. Diacylglycerol is a second messenger that activates protein kinase C (PKC) (Section 14.1). When active, PKC and other Ser/Thr protein kinases are capable of phosphorylating IRS and reducing the ability of IRS to propagate the insulin signal. Saturated and trans unsaturated fatty acids may also activate kinases that block the insulin signal. Ceramide or its metabolites inhibit glucose uptake and glycogen synthesis, apparently by inhibiting PDK and Akt (p. 808). The result is a diet-induced insulin resistance (Figure 27.7).

Insulin resistance in muscle facilitates pancreatic failure

What is the effect of overnutrition on the pancreas? This question is important because a primary function of the pancreas is to respond to the presence of glucose in the blood by secreting insulin, a process referred to as *glucose-stimulated insulin secretion* (GSIS). Indeed, the β cell is a virtual insulin factory. Proinsulin mRNA, which encodes proinsulin, the precursor to insulin, constitutes 20% of the total mRNA in the pancreas. Proinsulin constitutes 50% of the total protein synthesized in the pancreas.

Glucose enters the β cells of the pancreas through the glucose transporter GLUT2. Recall that GLUT2 will allow glucose transport only when blood glucose is plentiful, ensuring that insulin is secreted only when glucose is abundant, such as after a meal (Section 16.2). The β cell metabolizes glucose to CO_2 and H_2O in the process of cellular respiration, generating ATP (Chapters 16, 17, and 18). The resulting increase in the ATP/ADP ratio closes an ATP-sensitive K^+ channel that, when open, allows potassium to flow out of the cell (Figure 27.8). The resulting alteration in the cellular ionic environment opens a Ca^{2+} channel. The influx of Ca^{2+} causes insulin-containing secretory vesicles to fuse with the cell membrane and release insulin into the blood. Thus, the increase in energy charge resulting from the metabolism of glucose has been translated by the membrane proteins into a physiological response—the secretion of insulin and the subsequent removal of glucose from the blood.

What aspect of β cell function ultimately fails as a result of overnutrition, causing the transition from insulin resistance to full-fledged type 2 diabetes? Recall that, under normal circumstances, the β cells of the

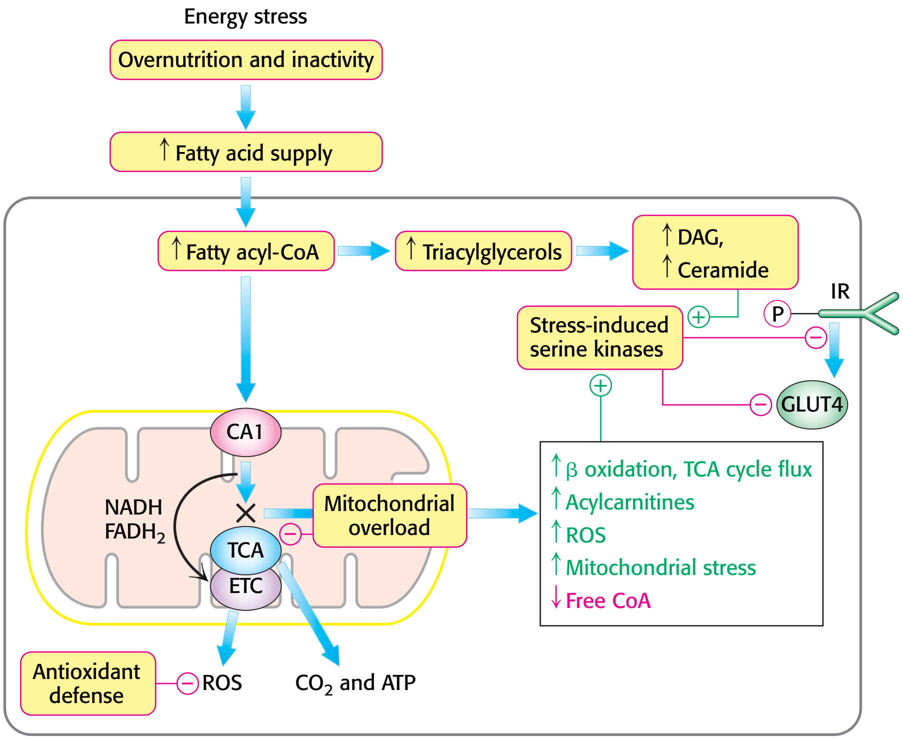
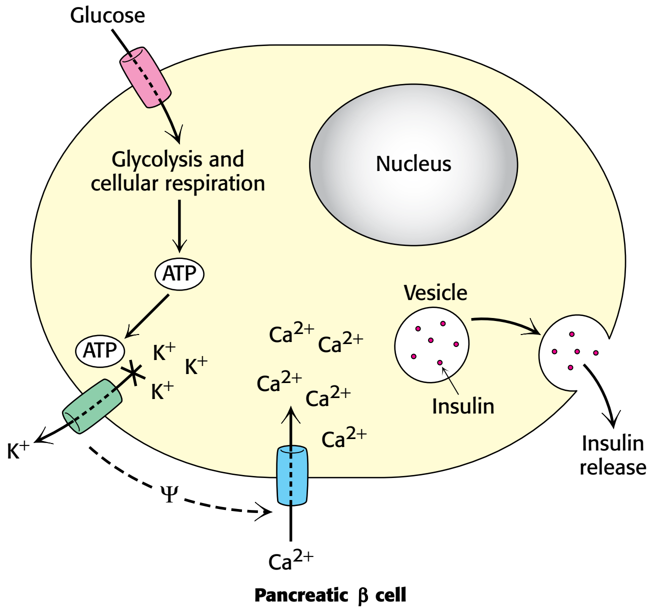

Energy stress
Overnutrition and inactivity
↑ Fatty acid supply
↑ Fatty acyl-CoA
↑ Triacylglycerols
↑ DAG,
↑ Ceramide
IR
P
Stress-induced serine kinases
GLUT4
CA1
NADH
FADH$_2$
Mitochondrial overload
TCA
ETC
↑ β oxidation, TCA cycle flux
↑ Acylcarnitines
↑ ROS
↑ Mitochondrial stress
↓ Free CoA
Antioxidant defense
ROS
CO$_2$ and ATP

FIGURE 27.7 Excess fat in peripheral tissues can result in insulin insensitivity. Excess fat accumulation in peripheral tissues, most notably muscle, can disrupt some signal-transduction pathways and inappropriately activate others. In particular, diacylglycerols and ceramide activate stress-induced pathways that interfere with insulin signaling, resulting in insulin resistance. (Abbreviations: DAG, diacylglycerol; TGs, triacylglycerols; ROS, reactive oxygen species; CA1, carnitine acyltransferase 1; GLUT4, glucose transporter 4; IR, insulin receptor; ETC, electron-transport chain; TCA, tricarboxylic acid cycle.)

pancreas synthesize large amounts of proinsulin. The proinsulin folds into its three-dimensional structure in the endoplasmic reticulum, is processed to insulin, and is subsequently packaged into vesicles for secretion. As insulin resistance develops in the muscle, the β cells respond by synthesizing yet more insulin in a futile attempt to drive insulin action. The ability of the endoplasmic reticulum to process all of the proinsulin and insulin becomes compromised, a condition known as *endoplasmic reticulum* (ER) *stress,* and unfolded or misfolded proteins accumulate. ER stress initiates a signal pathway called the *unfolded protein response* (UPR), a pathway intended to save the cell. UPR consists of several steps. First, general protein synthesis is inhibited so as to prevent more proteins from entering the ER. Second, chaperone synthesis is stimulated. Recall that chaperones are proteins that assist the folding of other proteins (Section 2.6). Third, misfolded proteins are removed from the ER and are subsequently delivered to the proteasome for destruction. Finally, if the described response fails to alleviate the ER stress, programmed cell death is triggered, which ultimately leads to β cell death and full-fledged type 2 diabetes.

What are the treatments for type 2 diabetes? Most are behavioral in nature. Diabetics are advised to count calories, making sure that energy intake does not exceed energy output; to consume a diet rich in vegetables, fruits, and grains; and to get plenty of aerobic exercise. Note that these guidelines are the same as those for healthy living, even for

FIGURE 27.8 Insulin release is regulated by ATP. The metabolism of glucose by glycolysis and cellular respiration increases the concentration of ATP, which causes an ATP-sensitive potassium channel to close. The closure of this channel alters the charge across the membrane (ψ) and causes a calcium channel to open. The influx of calcium causes insulin-containing granules to fuse with the plasma membrane, releasing insulin into the blood.


Glucose
Glycolysis and cellular respiration
Nucleus
ATP
Vesicle
ATP
K$^+$
K$^+$
K$^+$
K$^+$
Ca^{2+}
Ca^{2+}
Ca^{2+}
Ca^{2+}
Ca^{2+}
Insulin
Insulin release
Ψ
Ca^{2+}
Pancreatic β cell

those not suffering from type 2 diabetes. Treatments specific for type 2 diabetes include the monitoring of blood-glucose concentration so that it is within the target range (normal is 3.6 to 6.1 mM). For those who are not able to maintain proper glucose levels with the behaviors described herein, drug treatments are required. The administration of insulin may be necessary upon pancreatic failure, and treatment with metformin (Glucophage), which activates AMPK, may be effective. AMPK promotes the oxidation of fats while inhibiting fat synthesis and storage. It also stimulates glucose uptake and storage by muscle while inhibiting gluconeogenesis in the liver.

Metabolic derangements in type 1 diabetes result from insulin insufficiency and glucagon excess

We now turn to the more-straightforward type 1 diabetes. In type 1 diabetes, insulin production is insufficient because of autoimmune destruction of the β cells of the pancreas. Consequently, the glucagon/insulin ratio is at higher-than-normal levels. In essence, the diabetic person is in biochemical fasting mode despite a high concentration of blood glucose. Because insulin is deficient, *the entry of glucose into adipose and muscle cells is impaired*. The liver becomes stuck in a gluconeogenic and ketogenic state. The gluconeogenic state is characterized by excessive production of glucose. The excessive level of glucagon relative to that of insulin leads to a decrease in the amount of fructose 2,6-bisphosphate (F-2, 6-BP), which stimulates glycolysis and inhibits gluconeogenesis in the liver. Hence, glycolysis is inhibited and gluconeogenesis is stimulated because of the opposite effects of F-2, 6-BP on phosphofructokinase and fructose-1, 6-bisphosphatase (Section 16.4; Figure 27.9). Essentially, the cells' response to a lack of insulin amplifies the amount of glucose in the blood. The high glucagon/insulin ratio in diabetes also promotes glycogen breakdown. Hence, *an excessive amount of glucose is produced by the liver and released into the blood*. Glucose is excreted in the urine (hence the name *mellitus*) when its concentration in the blood exceeds the reabsorptive capacity of the renal tubules. Water accompanies the excreted glucose, and so an untreated diabetic in the acute phase of the disease is hungry and thirsty.

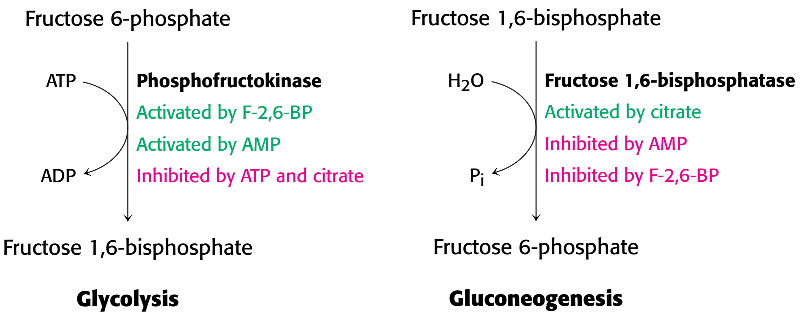


FIGURE 27.9 **Regulation of glycolysis and gluconeogenesis.** Phosphofructokinase is the key enzyme in the regulation of glycolysis, whereas fructose 1,6-bisphosphatase is the principal enzyme controlling the rate of gluconeogenesis. Note the reciprocal relation between the pathways and the signal molecules.

Because carbohydrate utilization is impaired, a lack of insulin leads to the uncontrolled breakdown of lipids and proteins, resulting in the ketogenic state. Large amounts of acetyl CoA are then produced by β oxidation. However, much of the acetyl CoA cannot enter the citric acid cycle, because there is insufficient oxaloacetate for the condensation step. Recall that mammals can synthesize oxaloacetate from pyruvate, a product of glycolysis, but not from acetyl CoA; instead, they generate ketone bodies (Section 22.3). *A striking feature of diabetes is the shift in fuel usage from carbohydrates to fats; glucose, more abundant than ever, is spurned.* In high concentrations, ketone bodies overwhelm the kidney's capacity to maintain acid–base balance. The untreated diabetic can go into a coma because of a lowered blood-pH level and dehydration. Interestingly, diabetic ketosis is rarely a problem in type 2 diabetes because insulin is active enough to prevent excessive lipolysis in liver and adipose tissue.

What is the treatment for type 1 diabetes? Many of the behaviors recommended for type 2 diabetes apply to type 1: watching calories, exercising, and eating a healthy diet. Likewise, blood-glucose levels must be monitored. Insulin treatments are required for survival.

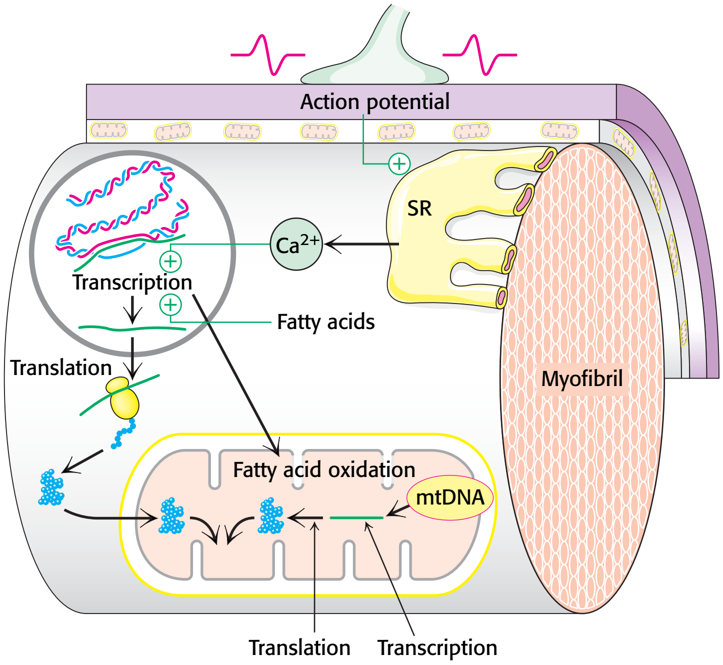
27.4 Exercise Beneficially Alters the Biochemistry of Cells

Skeletal muscle accounts for ≈40% of total body mass and ≈35% of resting metabolic activity. Moreover, skeletal muscle is the largest target tissue for insulin. Given its biochemical prominence, increased muscle activity—exercise—coupled with a healthy diet, is one of the most effective treatments for diabetes as well as a host of other pathological conditions. These include coronary disease, hypertension, depression, and a variety of cancers. With regard to diabetes, exercise increases the insulin sensitivity of people who are insulin resistant or type 2 diabetics. What is the basis of this beneficial effect?

Mitochondrial biogenesis is stimulated by muscular activity

When muscle is stimulated to contract during exercise by receiving nerve impulses from motor neurons, calcium is released from the sarcoplasmic reticulum. Calcium induces muscle contraction, as will be discussed in Chapter 35. Recall that calcium is also a potent second messenger and frequently works in association with the calcium-binding protein calmodulin (Section 14.1). In its capacity as a second messenger, calcium stimulates various calcium-dependent enzymes, such as calmodulin-dependent protein kinase. The calcium-dependent enzymes, as well as AMPK, subsequently activate particular transcription-factor complexes. As we will see in Chapters 29 and 31, transcription factors are proteins that control gene expression. Two patterns of gene expression, in particular, change in response to regular exercise (Figure 27.10). Regular exercise enhances the production of proteins required for fatty acid metabolism, such as the enzymes of β oxidation. Interestingly, fatty acids themselves function as signal molecules to activate the transcription of enzymes of fatty acid metabolism. Additionally, another set of transcription factors activated by the calcium signal cascade initiates a signal pathway that increases mitochondrial biogenesis. In concert, *the increase in fatty acid oxidizing capability and additional mitochondria allow for the efficient metabolism of fatty acids.* Because an excess of fatty acids results in insulin resistance, as already discussed, efficient metabolism of fatty acids results in *an increase in insulin sensitivity.* Indeed, muscles of well-trained athletes may contain high concentrations of triacylglycerols and still maintain exquisite sensitivity to insulin. These alterations of muscle biochemistry are some of the molecular manifestations of the training effect of exercise.

FIGURE 27.10 Exercise results in mitochondrial biogenesis and enhanced fat metabolism. An action potential causes Ca^{2+} release from the sarcoplasmic reticulum (SR), the muscle-cell equivalent of the endoplasmic reticulum. The Ca^{2+}, in addition to instigating muscle contraction, activates nuclear transcription factors that stimulate the expression of specific genes. The products of these genes, in conjunction with the products of mitochondrial genes, are responsible for mitochondrial biogenesis. Fatty acids activate a different set of genes that increase the fatty acid oxidation capability of mitochondria. [Information from D. A. Hood, *J Appl. Physiol.* 90:1137–1157, 2001, Fig. 2.]



Fuel choice during exercise is determined by the intensity and duration of activity

In keeping with our theme of energy use under different physiological conditions, we now examine how fuels are used in different types of exercise. The fuels used in anaerobic exercises—sprinting, for example—differ from those used in aerobic exercises—such as distance running. The selection of fuels during these different forms of exercise illustrates many important facets of energy transduction and metabolic integration. ATP directly powers myosin, the protein immediately responsible for converting chemical energy into movement (Chapter 35). However, the amount of ATP in muscle is small. Hence, the power output and, in turn, the velocity of running depend on the rate of ATP production from other fuels. As shown in Table 27.3, *creatine phosphate* (phosphocreatine) can swiftly

TABLE 27.3 Fuel sources for muscle contraction

Fuel source	Maximal rate of ATP production (mmol s^{-1})	Total ~P available (mmol)
Muscle ATP		223
Creatine phosphate	73.3	446
Conversion of muscle glycogen into lactate	39.1	6,700
Conversion of muscle glycogen into CO_2	16.7	84,000
Conversion of liver glycogen into CO_2	6.2	19,000
Conversion of adipose-tissue fatty acids into CO_2	6.7	4,000,000

Note: Fuels stored are estimated for a 70-kg person having a muscle mass of 28 kg.
Data from E. Hultman and R. C. Harris. In *Principles of Exercise Biochemistry,* edited by J. R. Poortmans (Karger, 2004), pp. 78–119.

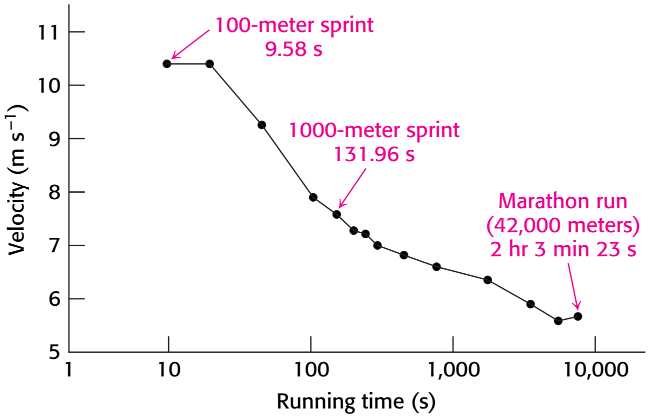
transfer its high-potential phosphoryl group to ADP to generate ATP. However, the amount of creatine phosphate, like that of ATP itself, is limited. Creatine phosphate and ATP can power intense muscle contraction for 5 to 6 s. Maximum speed in a sprint can thus be maintained for only 5 to 6 s (Figure 15.7). Thus, the winner in a 100-meter sprint is the runner who both achieves the highest initial velocity and then slows down the least.

During a ~10 second sprint, the ATP level in muscle drops from 5.2 to 3.7 mM, and that of creatine phosphate decreases from 9.1 to 2.6 mM. Anaerobic glycolysis provides fuel to make up for the loss of ATP and creatine phosphate. *A 100-meter sprint is powered by stored ATP, creatine phosphate, and the anaerobic glycolysis of muscle glycogen.* The conversion of muscle glycogen into lactate can generate a good deal more ATP, but the rate is slower than that of phosphoryl-group transfer from creatine phosphate. Because of anaerobic glycolysis, the blood-lactate level is elevated from 1.6 to 8.3 mM. The release of H^+ from the intensely active muscle concomitantly lowers the blood pH from 7.42 to 7.24. Sprinting is powered by fast twitch muscle fiber (type IIb) specialized for anaerobic glycolysis. Recall that one of the effects of high-intensity training is to increase the amount of lactate transporters in the membranes of slow twitch fibers (type 1), which remove lactate from the blood and slow the fall in blood pH (Section 16.4). Nonetheless, the pace of a 100-meter sprint cannot be sustained in a 1000-meter run (~132 s) for two reasons. First, creatine phosphate is consumed within a few seconds. Second, the lactate produced would cause acidosis. Thus, alternative fuel sources are needed.

Let us examine the fuel use of a middle-distance runner. The complete oxidation of muscle glycogen to CO_2 by aerobic respiration substantially increases the energy available to power the 1000-meter run, but this aerobic process is a good deal slower than anaerobic glycolysis. Because ATP is produced more slowly by oxidative phosphorylation than by glycolysis (Table 27.3), the runner's pace is necessarily slower than in a 100-meter sprint. The championship velocity for the 1000-meter run is about 7.6 m s^{-1}, compared with approximately 10.4 m s^{-1} for the 100-meter event (Figure 27.11).

The running of a marathon (26 miles 385 yards, or 42,200 meters) requires a different selection of fuels and is characterized by cooperation (depending on the capabilities of the runner) between muscle, liver, and adipose tissue. Liver glycogen complements muscle glycogen as an energy store that can be tapped. Elite marathoners can use the aerobic combustion of glucose to power the entire race, provided they consume fuels during the race. However, for most of us, the total body glycogen stores

FIGURE 27.11 Dependency of the velocity of running on the duration of the race. The values shown are world track records. [Data from trackandfieldnews.com.]



(103 mol of ATP at best) are insufficient to provide the 150 mol of ATP needed for this grueling event. Much larger quantities of ATP can be obtained by the oxidation of fatty acids derived from the breakdown of *fat in adipose tissue,* but the maximal rate of ATP generation is slower yet than that of glycogen oxidation and is more than 10-fold slower than that with creatine phosphate. Thus, *ATP is generated much more slowly from high-capacity stores than from limited ones*, accounting for the different velocities of anaerobic and aerobic events. Fats are rapidly consumed in activities such as distance running, explaining why extended aerobic exercise is beneficial for people who are insulin resistant.

Is it possible to determine the contribution of each fuel as a function of exercise intensity? The percentage contribution of each fuel can be measured with the use of a respirometer, which measures the respiratory quotient (RQ), the ratio of CO_2 produced to O_2 consumed. Consider the complete combustion of glucose:

$$\underset{\text{Glucose}}{C_6H_{12}O_6} + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O$$

The RQ for glucose is 1. Now consider the oxidation of a typical fatty acid, palmitate:

$$\underset{\text{Palmitate}}{C_{16}H_{32}O_2} + 23\,O_2 \longrightarrow 16\,CO_2 + 16\,H_2O$$

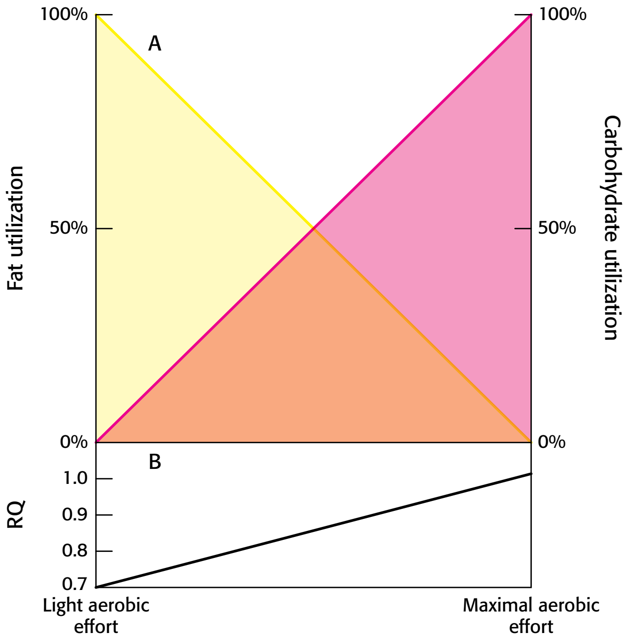
The RQ for palmitate oxidation is 0.7. Thus, as aerobic exercise intensity increases, the RQ will rise from 0.7 (only fats are used as fuel) to 1.0 (only glucose is used as fuel). Between these values, a mixture of fuels is used (Figure 27.12).

What is the optimal mix of fuels for use during a marathon? As suggested above, this is a complex question that varies with the athlete and level of training. Studies have shown that, when muscle glycogen has been depleted, the power output of the muscle falls to approximately 50% of maximum. Power output decreases despite the fact that ample supplies of fat are available, suggesting that fats can supply only about 50% of maximal aerobic effort. Indeed, depletion of glycogen stores during a race is referred to as "bonking" or "hitting the wall," with the result that the athlete must greatly reduce the pace.

How is an optimal mix of these fuels achieved? *A low blood-sugar level leads to a high glucagon/insulin ratio, which in turn mobilizes fatty acids from adipose tissue.* Fatty acids readily enter muscle, where they are degraded by β oxidation to acetyl CoA and then to CO_2. The elevated acetyl CoA level decreases the activity of the pyruvate dehydrogenase complex to block the conversion of pyruvate into acetyl CoA. Hence, fatty acid oxidation decreases the funneling of glucose into the citric acid cycle and oxidative phosphorylation. Glucose is spared so that just enough remains available at the end of the marathon to increase the pace as the finish line draws near. The simultaneous use of both fuels gives a higher mean velocity than would be attained if glycogen were totally consumed before the start of fatty acid oxidation. It is important to bear in mind that fuel utilization is only one of many factors that determine running ability.

If carbohydrate-rich meals are consumed after glycogen depletion, glycogen stores are rapidly restored. In addition, glycogen synthesis continues during the consumption of carbohydrate-rich meals, increasing glycogen stores far above normal. This phenomenon is called "super compensation" or, more commonly, carbo-loading.

FIGURE 27.12 An idealized representation of fuels use as a function of aerobic exercise intensity. (A) With increased exercise intensity, the use of fats as fuels falls as the utilization of glucose increases. (B) The respiratory quotient (RQ) measures the alteration in fuel use.



27.5 Food Intake and Starvation Induce Metabolic Changes

Thus far, we have been considering metabolism in the context of excess consumption of calories, as in obesity, or extreme caloric needs, as in exercise. We now look at the opposite physiological condition—lack of calories.

The starved–fed cycle is the physiological response to a fast

We begin with a physiological condition called the *starved–fed cycle*, which we all experience in the hours after an evening meal and through the night's fast. This nightly starved–fed cycle has three stages: the well-fed state after a meal, the early fasting during the night, and the refed state after breakfast. A major goal of the many biochemical alterations in this period is to maintain *glucose homeostasis*—that is, a constant blood-glucose concentration. Maintaining glucose homeostasis is crucial because glucose is normally the only fuel source for the brain. As discussed earlier, the major defect in diabetes is the inability to perform this vital task. The two primary signals regulating the starved–fed cycle are insulin and glucagon.

1. *The well-fed, or postprandial, state.* After we consume and digest an evening meal, glucose and amino acids are transported from the intestine to the blood. The dietary lipids are packaged into chylomicrons and transported to the blood by the lymphatic system. This fed condition leads to the secretion of insulin, which signals the fed state. Insulin stimulates glycogen synthesis in both muscle and the liver, suppresses gluconeogenesis by the liver, and stimulates protein synthesis. Insulin also accelerates glycolysis in the liver, which in turn increases the synthesis of fatty acids.

The liver helps to limit the amount of glucose in the blood during times of plenty by storing it as glycogen so as to be able to release glucose in times of scarcity. How is the excess blood glucose present after a meal removed? The liver is able to trap large quantities of glucose because it possesses an isozyme of hexokinase called *glucokinase*, which converts glucose into glucose 6-phosphate, which cannot be transported out of the cell. Recall that glucokinase has a high K_M value and is thus active only when blood-glucose concentration is high (Section 16.2). Furthermore, glucokinase is not inhibited by glucose 6-phosphate as hexokinase is. Consequently, *the liver forms glucose 6-phosphate more rapidly as the blood-glucose concentration rises. The increase in glucose 6-phosphate, which activates glycogen synthase, coupled with insulin action leads to a buildup of glycogen stores*. The hormonal effects on glycogen synthesis and storage are reinforced by a direct action of glucose itself. *Phosphorylase* a *is a glucose sensor in addition to being the enzyme that cleaves glycogen*. When the glucose level is high, the binding of glucose to phosphorylase *a* renders the enzyme susceptible to the action of a phosphatase that converts it into phosphorylase *b*, which does not readily degrade glycogen (Figure 21.24). Thus, *glucose allosterically shifts the glycogen system from a degradative to a synthetic mode*.

The high insulin level in the fed state also promotes *the entry of glucose into muscle and adipose tissue*. Insulin stimulates the synthesis of glycogen by muscle as well as by the liver. The entry of glucose into adipose tissue provides glycerol 3-phosphate for the synthesis of triacylglycerols. The action of insulin also extends to amino acid and protein metabolism. Insulin promotes the uptake of branched-chain amino acids (valine, leucine, and isoleucine) by muscle. Indeed, insulin has a general stimulating effect on protein synthesis, which favors a building up of muscle protein. In addition, it inhibits the intracellular degradation of proteins.

2. *The early fasting, or postabsorptive, state.* The blood-glucose concentration begins to drop several hours after a meal, leading to a decrease in insulin secretion and a rise in glucagon secretion by the α cells of the pancreas. The regulation of glucagon secretion is poorly understood, but when glucose is abundant, β cells inhibit glucagon secretion. When glucose concentration falls, the inhibition is relieved and glucagon is secreted by the α cells of the pancreas. Just as insulin signals the fed state, glucagon signals the starved state, serving to mobilize glycogen stores when there is no dietary intake of glucose. *The main target organ of glucagon is the liver.* Glucagon stimulates glycogen breakdown and inhibits glycogen synthesis by triggering the cyclic AMP cascade leading to the phosphorylation and activation of phosphorylase and the inhibition of glycogen synthase (Section 21.5). Glucagon also inhibits fatty acid synthesis by diminishing the production of pyruvate and by maintaining the inactive phosphorylated state of acetyl CoA carboxylase. In addition, glucagon stimulates gluconeogenesis in the liver and blocks glycolysis by lowering the level of F-2, 6-BP (Figure 27.9).

All known actions of glucagon are mediated by protein kinases that are activated by cyclic AMP. The activation of the cyclic AMP cascade results in a higher level of phosphorylase *a* activity and a lower level of glycogen synthase *a* activity. Glucagon's effect on this cascade is reinforced by the low concentration of glucose in the blood. The diminished binding of glucose to phosphorylase *a* makes the enzyme less susceptible to the hydrolytic action of the phosphatase. Instead, the phosphatase remains bound to phosphorylase *a*, and so the synthase stays in the inactive phosphorylated form. Consequently, there is a rapid mobilization of glycogen.

The large amount of glucose formed by the hydrolysis of glucose 6-phosphate derived from glycogen is then released from the liver into the blood. The entry of glucose into muscle and adipose tissue decreases in response to a low insulin level. The diminished utilization of glucose by muscle and adipose tissue also contributes to the maintenance of the blood-glucose concentration. The net result of these actions of glucagon is to *markedly increase the release of glucose by the liver*. Both muscle and the liver use fatty acids as fuel when the blood-glucose concentration drops, saving the glucose for use by the brain and red blood cells. Thus, *the blood-glucose concentration is kept at or above 4.4 mM by three major factors*: (1) *the mobilization of glycogen and the release of glucose by the liver,* (2) *the release of fatty acids by adipose tissue, and* (3) *the shift in the fuel used from glucose to fatty acids by muscle and the liver*.

What is the result of the depletion of the liver's glycogen stores? Gluconeogenesis from lactate and alanine continues, but this process merely replaces glucose that had already been converted into lactate and alanine by tissues such as muscle and red blood cells. Moreover, the brain oxidizes glucose completely to CO_2 and H_2O. Thus, for the net synthesis of glucose to take place, another source of carbon is required. Glycerol released from adipose tissue on lipolysis provides some of the carbon atoms, with the remaining carbon atoms coming from the hydrolysis of muscle proteins.

3. *The refed state.* What are the biochemical responses to a hearty breakfast? Fat is processed exactly as it is processed in the normal fed state. However, that is not the case for glucose. The liver does not initially absorb glucose from the blood, but, instead, leaves it for the other tissues. Moreover, the liver remains in a gluconeogenic mode. Now, however, the newly synthesized glucose is used to replenish the liver's glycogen stores. As the blood-glucose concentration continues to rise, the liver completes the replenishment of its glycogen stores and begins to process the remaining excess glucose for fatty acid synthesis.

TABLE 27.4 Fuel reserves in a typical 70-kg man

	Available energy in kilojoules (kcal)		
Organ	Glucose or glycogen	Triacylglycerols	Mobilizable proteins
Blood	250 (60)	20 (45)	0 (0)
Liver	1700 (400)	2000 (450)	1700 (400)
Brain	30 (8)	0 (0)	0 (0)
Muscle	5000 (1200)	2000 (450)	100,000 (24,000)
Adipose tissue	330 (80)	560,000 (135,000)	170 (40)

Data from G. F. Cahill, Jr., *Clin. Endocrinol. Metab.* 5:398, 1976.

Metabolic adaptations in prolonged starvation minimize protein degradation

Earlier, we considered the metabolic results of overnutrition, a condition becoming all too common in prosperous nations. Let us now examine the opposite extreme. What are the adaptations if fasting is prolonged to the point of starvation, a circumstance affecting nearly a billion people worldwide? A typical well-nourished 70-kg man has fuel reserves totaling about 670,000 kJ (161,000 kcal; Table 27.4). The energy need for a 24-hour period ranges from about 6700 kJ (1600 kcal) to 25,000 kJ (6000 kcal), depending on the extent of activity. Thus, stored fuels suffice to meet caloric needs in starvation for 1 to 3 months. However, the carbohydrate reserves are exhausted in only a day.

Even under starvation conditions, the blood-glucose concentration must be maintained above 2.2 mM. *The first priority of metabolism in starvation is to provide sufficient glucose to the brain and other tissues (such as red blood cells) that are absolutely dependent on this fuel.* However, precursors of glucose are not abundant. Most energy is stored in the fatty acyl moieties of triacylglycerols. Moreover, recall that fatty acids cannot be converted into glucose, because acetyl CoA resulting from fatty acid breakdown cannot be transformed into pyruvate (Section 22.3). The glycerol moiety of triacylglycerol can be converted into glucose, but only a limited amount is available. The only other potential source of glucose is the carbon skeletons of amino acids derived from the breakdown of proteins. However, proteins are not stored, and so any breakdown will necessitate a loss of function. Thus, *the second priority of metabolism in starvation is to preserve protein, which is accomplished by shifting the fuel being used from glucose to fatty acids and ketone bodies* (Figure 27.13).

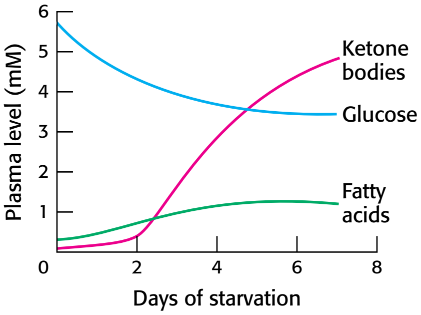


FIGURE 27.13 Fuel choice during starvation. The plasma levels of fatty acids and ketone bodies increase in starvation, whereas that of glucose decreases.

The metabolic changes on the first day of starvation are like those after an overnight fast. The low blood-sugar level leads to decreased secretion of insulin and increased secretion of glucagon. *The dominant metabolic processes are the mobilization of triacylglycerols in adipose tissue and gluconeogenesis by the liver.* The liver obtains energy for its own needs by oxidizing fatty acids released from adipose tissue. The concentrations of acetyl CoA and citrate consequently increase, which switches off glycolysis. The uptake of glucose by muscle is markedly diminished because of the low insulin level, whereas fatty acids enter freely. Consequently, *muscle uses no glucose and relies exclusively on fatty acids for fuel*. The β oxidation of fatty acids by muscle halts the conversion of pyruvate into acetyl CoA, because acetyl CoA stimulates the phosphorylation of the pyruvate dehydrogenase complex, which renders it inactive (Section 17.3). Hence, any available pyruvate, lactate, and alanine are exported to the liver for conversion into glucose. Glycerol derived from the cleavage of triacylglycerols is another raw material for the synthesis of glucose by the liver.

Proteolysis also provides carbon skeletons for gluconeogenesis. During starvation, degraded proteins are not replenished and serve as carbon

TABLE 27.5 Fuel metabolism in starvation

	Amount formed or consumed in 24 hours (grams)	
Fuel exchanges and consumption	3d day	40th day
Fuel use by the brain		
Glucose	100	40
Ketone bodies	50	100
All other use of glucose	50	40
Fuel mobilization		
Adipose-tissue lipolysis	180	180
Muscle-protein degradation	75	20
Fuel output of the liver		
Glucose	150	80
Ketone bodies	150	150

sources for glucose synthesis. Initial sources of protein are those that turn over rapidly, such as proteins of the intestinal epithelium and the secretions of the pancreas. Proteolysis of muscle protein provides some of the three-carbon precursors of glucose. However, survival for most animals depends on being able to move rapidly, which requires a large muscle mass, and so muscle loss must be minimized.

How is the loss of muscle curtailed? After about 3 days of starvation, the liver forms large amounts of the ketone bodies, acetoacetate and D-3-hydroxybutyrate (Figure 27.14). Their synthesis from acetyl CoA increases markedly because the citric acid cycle is unable to oxidize all the acetyl units generated by the degradation of fatty acids. Gluconeogenesis depletes the supply of oxaloacetate, which is essential for the entry of acetyl CoA into the citric acid cycle. Consequently, the liver produces large quantities of ketone bodies, which are released into the blood. At this time, *the brain begins to consume significant amounts of acetoacetate in place of glucose*. After 3 days of starvation, about a quarter of the energy needs of the brain are met by ketone bodies (Table 27.5). The heart also uses ketone bodies as fuel.

After several weeks of starvation, ketone bodies become the major fuel of the brain. Acetoacetate is activated by the transfer of CoA from succinyl CoA to give acetoacetyl CoA (Figure 27.15). Cleavage by thiolase then yields two molecules of acetyl CoA, which enter the citric acid cycle. In essence, *ketone bodies are equivalents of fatty acids that are an accessible fuel source for the brain*. Only 40 g of glucose is then needed per day for the brain, compared with about 120 g in the first day of starvation. *The effective conversion of fatty acids into ketone bodies by the liver and their use by the brain markedly diminishes the need for glucose. Hence, less muscle is degraded than in the first days of starvation*. The breakdown of 20 g of muscle daily compared with 75 g early in starvation is most important for survival. A person's survival time is mainly determined by the size of the triacylglycerol depot.

What happens after depletion of the triacylglycerol stores? The only source of fuel that remains is protein. Protein degradation accelerates, and death inevitably results from a loss of heart, liver, or kidney function.

27.6 Ethanol Alters Energy Metabolism in the Liver

Ethanol has been a part of the human diet for centuries, partly because of its intoxicating effects and partly because alcoholic beverages provided a safe means of hydration when pure water was scarce. However, its consumption

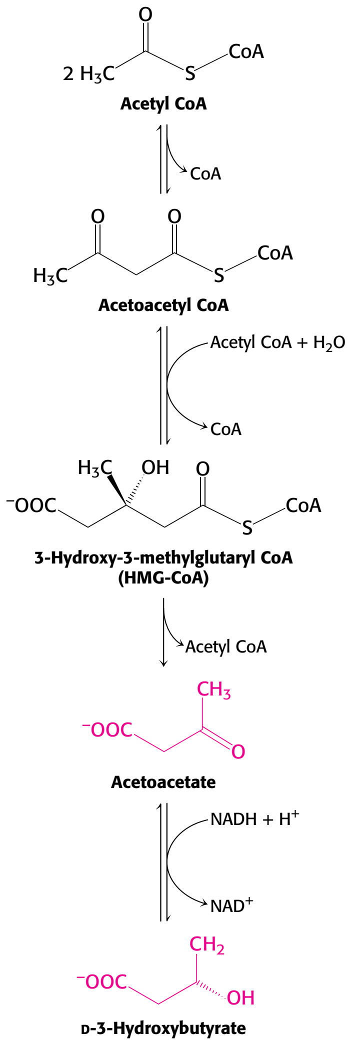


FIGURE 27.14 Synthesis of ketone bodies by the liver.

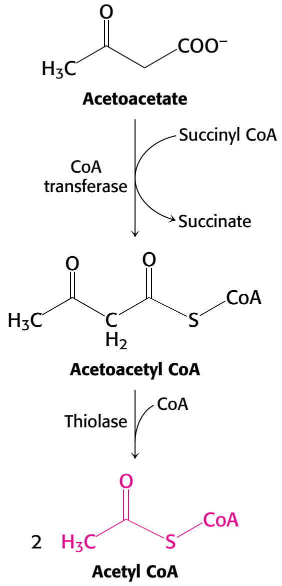


FIGURE 27.15 Entry of ketone bodies into the citric acid cycle.

in excess can result in a number of health problems, most notably liver damage. What is the biochemical basis of these health problems?

Ethanol metabolism leads to an excess of NADH

Ethanol cannot be excreted and must be metabolized, primarily by the liver. This metabolism is accomplished by two pathways. The first pathway comprises two steps. The first step, catalyzed by the enzyme *alcohol dehydrogenase,* takes place in the cytoplasm:

$$\underset{\textbf{Ethanol}}{\mathrm{CH_3CH_2OH}} + \mathrm{NAD^+} \xrightarrow[]{\text{Alcohol dehydrogenase}} \underset{\textbf{Acetaldehyde}}{\mathrm{CH_3CHO}} + \mathrm{NADH} + \mathrm{H^+}$$

The second step, catalyzed by *aldehyde dehydrogenase,* takes place in mitochondria:

$$\underset{\textbf{Acetaldehyde}}{\mathrm{CH_3CHO}} + \mathrm{NAD^+} + \mathrm{H_2O} \xrightarrow[]{\text{Aldehyde dehydrogenase}} \underset{\textbf{Acetate}}{\mathrm{CH_3COO^-}} + \mathrm{NADH} + \mathrm{H^+}$$

Note that *ethanol consumption leads to an accumulation of NADH*. This high concentration of NADH inhibits gluconeogenesis by preventing the oxidation of lactate to pyruvate. In fact, the high concentration of NADH will cause the reverse reaction to predominate, and lactate will accumulate. The consequences may be hypoglycemia and lactic acidosis.

The overabundance of NADH also inhibits fatty acid oxidation. An important metabolic purpose of fatty acid oxidation is to generate NADH for ATP generation by oxidative phosphorylation, but an alcohol consumer's NADH needs are met by ethanol metabolism. In fact, the excess NADH signals that conditions are right for fatty acid synthesis. Hence, triacylglycerols accumulate in the liver, leading to a condition known as "fatty liver" that is exacerbated in obese persons. The biochemical effects of ethanol consumption can be quite rapid. For instance, fat accumulates in the liver within a few days of moderate alcohol consumption. This accumulation is reversible with a decrease in alcohol intake.

The second pathway for ethanol metabolism uses cytochrome P450 enzymes. This means of ethanol metabolism is also called the *microsomal ethanol-oxidizing system* (MEOS). This cytochrome P450-dependent pathway (Section 26.4) generates acetaldehyde and subsequently acetate while oxidizing biosynthetic reducing power, NADPH, to $\mathrm{NADP^+}$. Because it uses oxygen, this pathway generates free radicals that damage tissues. Moreover, because the system consumes NADPH, the antioxidant glutathione cannot be regenerated (Section 20.5), exacerbating the oxidative stress.

What are the effects of the other metabolites of ethanol? Liver mitochondria can convert acetate into acetyl CoA in a reaction requiring ATP. The enzyme is the thiokinase that normally activates short-chain fatty acids.

$$\text{Acetate} + \text{coenzyme A} + \text{ATP} \longrightarrow \text{acetyl CoA} + \text{AMP} + \mathrm{PP_i}$$

$$\mathrm{PP_i} \longrightarrow 2\ \mathrm{P_i}$$

However, further processing of the acetyl CoA by the citric acid cycle is blocked, because NADH inhibits two important citric acid cycle regulatory enzymes—isocitrate dehydrogenase and α-ketoglutarate dehydrogenase. The accumulation of acetyl CoA has several consequences. First, ketone bodies will form and be released into the blood, aggravating the acidic condition already resulting from the high lactate concentration. The processing of the acetate in the liver becomes inefficient, leading to a buildup of acetaldehyde. This very reactive compound forms covalent bonds with many important functional groups in proteins, impairing protein function. If ethanol is consistently consumed at high levels, the acetaldehyde can significantly damage the liver, eventually leading to cell death.

Liver damage from excessive ethanol consumption occurs in three stages. The first stage is the aforementioned development of fatty liver. In the second stage—alcoholic hepatitis—groups of cells die and inflammation results. This stage can itself be fatal. In stage three—cirrhosis—fibrous structure and scar tissue are produced around the dead cells. Cirrhosis impairs many of the liver's biochemical functions. The cirrhotic liver is unable to convert ammonia into urea, and blood levels of ammonia rise. Ammonia is toxic to the nervous system and can cause coma and death. Cirrhosis of the liver arises in about 25% of alcoholics, and about 75% of all cases of liver cirrhosis are the result of alcoholism. Viral hepatitis is a non-alcoholic cause of liver cirrhosis.

Excess ethanol consumption disrupts vitamin metabolism

The adverse effects of ethanol are not limited to the metabolism of ethanol itself. Vitamin A (retinol) is converted into retinoic acid, an important signal molecule for growth and development in vertebrates, by the same dehydrogenases that metabolize ethanol. Consequently, this activation does not take place in the presence of ethanol, which acts as a competitive inhibitor. Moreover, the P450 enzymes induced by ethanol inactivate retinoic acid. These disruptions in the retinoic acid signaling pathway are believed to be responsible, at least in part, for fetal alcohol syndrome as well as the development of a variety of cancers.

The disruption of vitamin A metabolism is a direct result of the biochemical changes induced by excess ethanol consumption. Other disruptions in metabolism result from another common characteristic of alcoholics—malnutrition. Alcoholics will frequently drink instead of eating. A dramatic neurological disorder, referred to as *Wernicke–Korsakoff syndrome,* results from insufficient intake of the vitamin thiamine. Symptoms include mental confusion, unsteady gait, and lack of fine motor skills. The symptoms of Wernicke–Korsakoff syndrome are similar to those of beriberi (Section 17.4) because both conditions result from a lack of thiamine. Thiamine is converted into the coenzyme thiamine pyrophosphate, a key constituent of the pyruvate dehydrogenase complex. Recall that this complex links glycolysis with the citric acid cycle. Disruptions in the pyruvate dehydrogenase complex are most evident as neuromuscular disorders because the nervous system is normally dependent on glucose for energy generation.

Alcoholic scurvy is occasionally observed because of an insufficient ingestion of vitamin C. Vitamin C is required for the formation of stable collagen fibers. The symptoms of scurvy include skin lesions and blood-vessel fragility. Most notable are bleeding gums, the loss of teeth, and periodontal infections. Gums are especially sensitive to a lack of vitamin C because the collagen in gums turns over rapidly. What is the biochemical basis for scurvy? Vitamin C is required for the synthesis of 4-hydroxyproline, an amino acid necessary for collagen stability. To form this unusual amino acid, proline residues on the amino side of glycine residues in nascent collagen chains become hydroxylated. One oxygen atom from O_2 becomes attached to C-4 of proline while the other oxygen atom is taken up by α-ketoglutarate, which is converted into succinate (Figure 27.16). This reaction is catalyzed by *prolyl hydroxylase,* a *dioxygenase,* which requires an Fe^{2+} ion to activate O_2. The enzyme also converts α-ketoglutarate into succinate without hydroxylating proline. In this partial reaction, an oxidized iron complex is formed, which inactivates the enzyme. How is the active enzyme regenerated? *Ascorbate (vitamin C)* comes to the rescue by reducing the ferric ion of

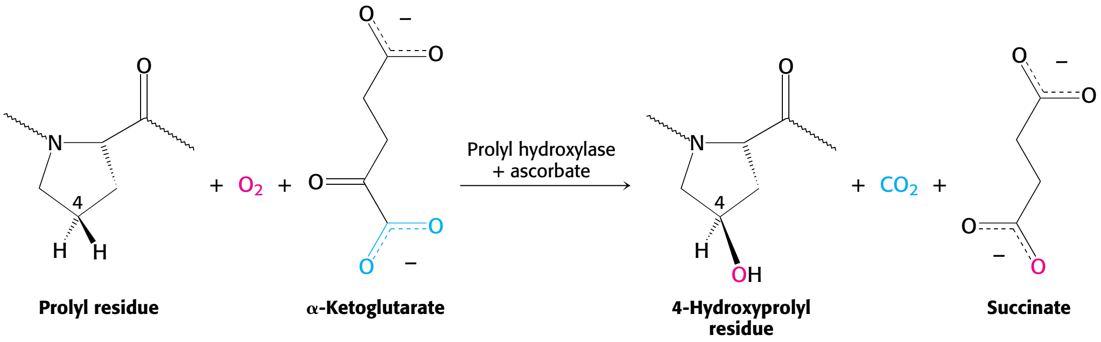


FIGURE 27.16 Formation of 4-hydroxyproline. Proline is hydroxylated at C-4 by the action of prolyl hydroxylase, an enzyme that activates molecular oxygen.

the inactivated enzyme. In the recovery process, ascorbate is oxidized to dehydroascorbic acid (Figure 27.17). Thus, ascorbate serves here as a specific *antioxidant*. Why does impaired hydroxylation have such devastating consequences? *Collagen synthesized in the absence of ascorbate is less stable than the normal protein*. Hydroxyproline stabilizes the collagen triple helix by forming interstrand hydrogen bonds. The abnormal fibers formed by insufficiently hydroxylated collagen account for the symptoms of scurvy.

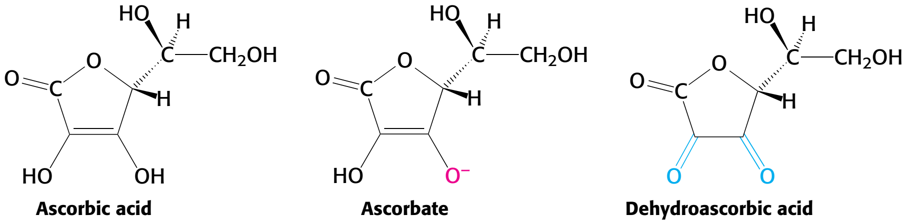


FIGURE 27.17 Forms of ascorbic acid (vitamin C). Ascorbate is the ionized form of vitamin C, and dehydroascorbic acid is the oxidized form of ascorbate.

SUMMARY

27.1 Caloric Homeostasis Is a Means of Regulating Body Weight

Many people are able to maintain a near-constant body weight throughout adult life. This ability is a demonstration of caloric homeostasis, a physiological condition in which energy needs match energy intake. When energy intake is greater than energy needs, weight gain results. In the developed world, obesity is at epidemic proportions and is implicated as a contributing factor in a host of pathological conditions.

27.2 The Brain Plays a Key Role in Caloric Homeostasis

Various signal molecules act on the brain to control appetite. Short-term signals such as CCK and GLP-1 relay satiety signals to the brain while eating is in progress. Long-term signals include leptin and insulin. Leptin, secreted by adipose tissue in direct proportion to adipose-tissue mass, is an indication of fat stores. Leptin inhibits eating. Insulin also works in the brain, signaling carbohydrate availability.

Leptin acts by binding to a receptor in brain neurons, which initiates signal-transduction pathways that reduce appetite. Obesity can develop in individuals with normal amounts of leptin and the leptin receptor, suggesting that such individuals are leptin resistant. Suppressors of cytokine signaling may inhibit leptin signaling, leading to leptin resistance and obesity.

27.3 Diabetes Is a Common Metabolic Disease Often Resulting from Obesity

Diabetes is the most common metabolic disease in the world. Type 1 diabetes results when insulin is absent due to autoimmune destruction of the β cells of the pancreas. Type 2 diabetes is characterized by normal or higher levels of insulin, but the target tissues of insulin, notably muscle, do not respond to the hormone, a condition called insulin resistance. Obesity is a significant predisposing factor for type 2 diabetes.

In muscle, excess fats accumulate in an obese individual. These fats are processed to second messengers that activate signal-transduction pathways that inhibit insulin signaling, leading to insulin resistance. Insulin resistance in target tissues ultimately leads to pancreatic β-cell failure. The pancreas tries to compensate for a lack of insulin action by synthesizing more insulin, resulting in ER stress and subsequent activation of apoptotic pathways that lead to β-cell death.

Type 1 diabetes is due to metabolic derangements resulting in an insufficiency of insulin and an excess of glucagon relative to a person's needs. The result is an elevated blood-glucose level, the mobilization of triacylglycerols, and excessive ketone-body formation. Accelerated ketone-body formation can lead to acidosis, coma, and death in untreated insulin-dependent diabetics.

27.4 Exercise Beneficially Alters the Biochemistry of Cells

Exercise is a useful prescription for insulin resistance and type 2 diabetes. Muscle activity stimulates mitochondrial biogenesis in a calcium-dependent manner. The increase in the number of mitochondria facilitates fatty acid oxidation in the muscle, resulting in increased insulin sensitivity.

Fuel choice in exercise is determined by the intensity and duration of the bout of exercise. Sprinting and marathon running are powered by different fuels to maximize power output. The 100-meter sprint is powered by stored ATP, creatine phosphate, and anaerobic glycolysis. In contrast, the oxidation of both muscle glycogen and fatty acids derived from adipose tissue is essential in the running of a marathon, a highly aerobic process.

27.5 Food Intake and Starvation Induce Metabolic Changes

Insulin signals the fed state; it stimulates the formation of glycogen and triacylglycerols and the synthesis of proteins. In contrast, glucagon signals a low blood-glucose level; it stimulates glycogen breakdown and gluconeogenesis by the liver and triacylglycerol hydrolysis by adipose tissue. After a meal, the rise in the blood-glucose level leads to an increased secretion of insulin and a decreased secretion of glucagon. Consequently, glycogen is synthesized in muscle and the liver. When the blood-glucose level drops several hours later, glucose is then formed by the degradation of glycogen and by the gluconeogenic pathway, and fatty acids are released by the hydrolysis of triacylglycerols. The liver and muscle then increasingly use fatty acids instead of glucose to meet their own energy needs so that glucose is conserved for use by the brain and the red blood cells.

The metabolic adaptations in starvation serve to minimize protein degradation. Large amounts of ketone bodies are formed by the liver from fatty acids and released into the blood within a few days after the onset of starvation. After several weeks of starvation, ketone bodies become the major fuel of the brain. The diminished need for glucose decreases the rate of muscle breakdown, and so the likelihood of survival is enhanced.

27.6 Ethanol Alters Energy Metabolism in the Liver

The oxidation of ethanol results in an unregulated overproduction of NADH, which has several consequences. A rise in the blood levels of lactic acid and ketone bodies causes a fall in blood pH, or acidosis. The liver is damaged because the excess NADH causes excessive fat formation as well as the generation of acetaldehyde, a reactive molecule. Severe liver damage can result.

KEY TERMS

caloric homeostasis (energy homeostasis) (p. 801)
cholecystokinin (CCK) (p. 804)
glucagon-like peptide 1 (GLP-1) (p. 805)
leptin (p. 805)
insulin (p. 805)
adiponectin (p. 806)
leptin resistance (p. 806)
type 1 diabetes (p. 808)
insulin resistance (p. 808)
type 2 diabetes (p. 808)
metabolic syndrome (p. 809)
endoplasmic reticulum (ER) stress (p. 811)
unfolded protein response (UPR) (p. 811)
starved–fed cycle (p. 816)
glucose homeostasis (p.816)

PROBLEMS

1. *Weight control.* As disturbing as the obesity epidemic is, an equally intriguing, almost amazing observation is that many people are able to maintain an approximately constant weight throughout adult life. A few simple calculations of a simplified situation illustrates how remarkable this feat is. Consider a 120-pound woman whose weight does not change significantly between the ages of 25 and 65. Let us say that the woman requires 8,400 kJ (2000 kcal) day^{-1}. For simplicity's sake, let us assume that the woman's diet consists predominantly of fatty acids derived from lipids. The energy density of fatty acids is 38 kJ (9 kcal) g^{-1}. How much food has she consumed over the 40 years in question?

2. *Spare tire.* Suppose that our test subject from problem 1 gained 55 pounds between the ages of 25 and 65 (a sadly common occurrence), and that her weight at 65 years of age is 175 pounds. Calculate how many excess calories she consumed per day to gain the 55 pounds over 40 years. Assume that our test subject is 5 feet 6 inches tall. What is her BMI? Would she be considered obese at 175 lbs?

3. *Depot fat.* Adipose tissue was once only considered a storage site for fat. Why is this view no longer considered correct?

4. *Balancing act.* What is meant by caloric homeostasis?

5. *Dynamic duo.* What are the key hormones responsible for maintaining caloric homeostasis?

6. *Dual roles.* What two biochemical roles does CCK play? GLP-1?

7. *Failure to communicate.* Leptin inhibits eating and is secreted in amounts in direct proportion to body fat. Moreover, obese people have normal amounts of leptin and leptin receptor. Why, then, do people become obese?

8. *Many signals.* Match the characteristic (a–i) with the appropriate hormone (1–6).

(a) Secreted by adipose tissue _____
(b) Stimulates liver gluconeogenesis _____
(c) GPCR pathway _____
(d) Satiety signal _____
(e) Enhances insulin secretion _____
(f) Secreted by the pancreas during a fast _____
(g) Secreted after a meal _____
(h) Stimulates glycogen synthesis _____
(i) Missing in type 1 diabetes _____

1. leptin
2. adiponectin
3. GLP-1
4. CCK
5. insulin
6. glucagon

9. *A key chemical.* What are the sources of glucose 6-phosphate in liver cells?

10. *Neither option is good.* Differentiate between type 1 and type 2 diabetes.

11. *Fighting diabetes.* Leptin is considered an "anti-diabetogenic" hormone. Explain.

12. *Metabolic energy and power.* The rate of energy expenditure of a typical 70-kg person at rest is about 70 watts (W), like that of a light bulb.

(a) Express this rate in kilojoules per second and in kilocalories per second.

(b) How many electrons flow through the mitochondrial electron-transport chain per second under these conditions?

(c) Estimate the corresponding rate of ATP production.

(d) The total ATP content of the body is about 50 g. Estimate how often an ATP molecule turns over in a person at rest.

13. *Respiratory quotient (RQ).* This classic metabolic index is defined as the volume of CO_2 released divided by the volume of O_2 consumed.

(a) Calculate the RQ values for the complete oxidation of glucose and of tripalmitoylglycerol.

(b) What do RQ measurements reveal about the contributions of different energy sources during intense exercise? (Assume that protein degradation is negligible.)

14. *Camel's hump.* Compare the H_2O yield from the complete oxidation of 1 g of glucose with that of 1 g of tripalmitoylglycerol. Relate these values to the evolutionary selection of the contents of a camel's hump.

15. *Hungry–nourished.* What is meant by the starved–fed cycle?

16. *Of course, too much is bad for you.* What are the primary means of processing ethanol?

17. *Started out with burgundy, but soon hit the harder stuff.* Describe the three stages of ethanol consumption that lead to liver damage and possibly death.

18. *The wages of sin.* How long does a person have to jog to offset the calories obtained from eating 10 macadamia nuts (75 kJ, or 18 kcal, per nut)? (Assume an incremental power consumption of 400 W.)

19. *Sweet hazard.* Ingesting large amounts of glucose before a marathon might seem to be a good way of increasing the fuel stores. However, experienced runners do not ingest glucose before a race. What is the biochemical reason for their avoidance of this potential fuel? (Hint: Consider the effect of glucose ingestion on the level of insulin.)

20. *Lipodystrophy.* Lipodystrophy is a condition in which an individual lacks adipose tissue. The muscles and liver from such individuals are insulin resistant, and both tissues accumulate large amounts of triacylglycerols (hyperlipidemia). The administration of leptin partly ameliorates this condition. What does it indicate about the relation of adipose tissue to insulin action?

21. *Therapeutic target.* What would be the effect of a mutation in the gene for PTP1B (protein tyrosine phosphatase 1B) that inactivated the enzyme in a person who has type 2 diabetes?

22. *An effect of diabetes.* Insulin-dependent diabetes is often accompanied by hypertriglyceridemia, which is an excess blood level of triacylglycerols in the form of very low density lipoproteins. Suggest a biochemical explanation.

23. *Sharing the wealth.* The hormone glucagon signifies the starved state, yet it inhibits glycolysis in the liver. How does this inhibition of an energy-production pathway benefit the organism?

24. *Compartmentation.* Glycolysis takes place in the cytoplasm, whereas fatty acid degradation takes place in mitochondria. What metabolic pathways depend on the interplay of reactions that take place in both compartments?

25. *Kwashiorkor.* The most common form of malnutrition in children in the world, kwashiorkor, is caused by a diet having ample calories but little protein. The high levels of carbohydrate result in high levels of insulin. What is the effect of high levels of insulin on

(a) lipid utilization?

(b) protein metabolism?

(c) Children suffering from kwashiorkor often have large distended bellies caused by water from the blood leaking into extracellular spaces. Suggest a biochemical basis for this condition.

26. *One for all, all for one.* How is the metabolism of the liver coordinated with that of skeletal muscle during strenuous exercise?

27. *A little help, please?* What is the advantage of converting pyruvate into lactate in skeletal muscle?

28. *Fuel choice.* What is the major fuel for resting muscle? What is the major fuel for muscle under strenuous work conditions?

29. *Hefty reimbursement.* Endurance athletes sometimes follow the exercise-and-diet plan described here: 7 days before an event, do exhaustive exercises so as to all but deplete glycogen stores. For the next 2 to 3 days, consume few carbohydrates and do moderate- to low-intensity exercises. Finally, 3 to 4 days before the event, consume a diet rich in carbohydrates. Explain the benefits of this regime.

30. *Oxygen deficit.* After light exercise, the oxygen consumed in recovery is approximately equal to the oxygen deficit, which is the amount of additional oxygen that would have been consumed had oxygen consumption reached steady state immediately. How is the oxygen consumed in recovery used?

31. *Excess post-exercise oxygen consumption.* The oxygen consumed after strenuous exercise stops is significantly greater than the oxygen deficit and is termed *excess post-exercise oxygen consumption* (EPOC). Why is so much more oxygen required after intense exercise?

32. *Psychotropic effects.* Ethanol is unusual in that it is freely soluble in both water and lipids. Thus, it has access to all regions of the highly vascularized brain. Although the molecular basis of ethanol action in the brain is not clear, ethanol evidently influences a number of neurotransmitter

receptors and ion channels. Suggest a biochemical explanation for the diverse effects of ethanol.

33. *Fiber type.* Skeletal muscle has several distinct fiber types. Type I is used primarily for aerobic activity, whereas type IIb is specialized for short, intense bursts of activity. How could you distinguish between these types of muscle fiber if you viewed them with an electron microscope?

34. *Tour de France.* Cyclists in the Tour de France (more than 2000 miles in 3 weeks) require about 836,000 kJ (200,000 kcal) of energy, or 41,840 kJ (10,000 kcal) day^{-1} (a resting male requires, 8368 kJ, or 2000 kcal, day^{-1}).

(a) With the assumptions that the energy yield of ATP is about 50.2 kJ (12 kcal) mol^{-1} and that ATP has a molecular weight of 503 g mol^{-1}, how much ATP would be expended by a Tour de France cyclist?

(b) Pure ATP can be purchased at a cost of approximately \$150 per gram. How much would it cost to power a cyclist through the Tour de France if the ATP had to be purchased?

35. *Responding to stress.* Why does it make good physiological sense that regular bouts of prolonged exercise will result in mitochondrial biogenesis?

36. *Burning fats.* How does the activation of AMP-activated protein kinase (AMPK) during aerobic exercise foster the switch to fatty acid oxidation for distance running?

37. *Saving protein.* What metabolic and hormonal changes account for decreased gluconeogenesis during the first several weeks of starvation in humans?

38. *Did I do that!?* Alcohol consumption on an empty stomach results in some interesting biochemical as well as embarrassing behavioral alterations. We will ignore the latter. Gluconeogenesis falls; there are increases in intracellular ratios of lactate to pyruvate, of glycerol 3-phosphate to dihydroxyacetone phosphate, of glutamate to α-ketoglutarate, and of D-3-hydroxybutyrate to acetoacetate. Hypoglycemia develops rapidly. Blood pH also falls. Alcohol consumption by a well-fed individual does not lead to hypoglycemia or alteration in blood pH.

(a) Why does ethanol consumption result in the altered ratios?

(b) Why does hypoglycemia and blood acidosis result in the hungry individual?

(c) Why does the well-fed person not experience hypoglycemia?

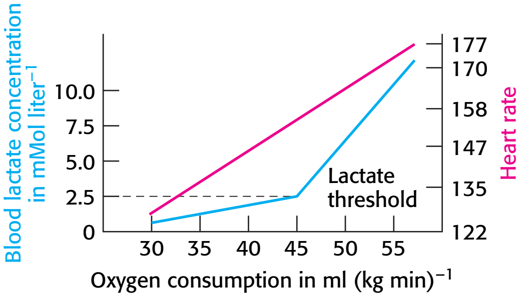
39. *Skip the candy bar.* Humans can survive longer by total fasting than on a diet consisting entirely of small amounts of carbohydrates. What is going on?

40. *Too much of a good thing.* What is the relation between fatty acid oxidation and insulin resistance in the muscle?

41. *Aneurin? Really?* Why are the symptoms of beriberi similar to those of Wernicke–Korsakoff syndrome?

Data Interpretation Problem

42. *Lactate threshold.* The graph shows the relation between blood-lactate levels, oxygen consumption, and heart rate during exercise of increasing intensity. The values for oxygen consumption and heart rate are indicators of the degree of exertion.



(a) Why is some lactate produced even when exercise is moderate?

(b) Biochemically, what is taking place when the lactate concentration begins to rise rapidly, a point called the lactate threshold?

(c) Endurance athletes will sometimes measure blood-lactate levels during training so that they know their lactate threshold. Then, during events, they will race just at or below their lactate threshold until the late stages of the race. Biochemically, why is this practice wise?

(d) Training can increase the lactate threshold. Explain.

Chapter Integration Problems

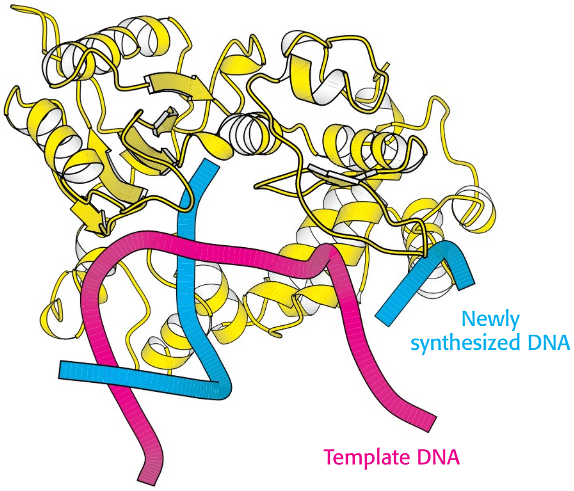
43. *Feeling the burn.* Under certain circumstances, the respiratory quotient (RQ value) for an athlete exercising intensely can rise above one. How is this possible?

44. *So many channels. Like cable TV.* Describe the role of ion channels in insulin secretion by the β cells of the pancreas.

CHAPTER

28

DNA Replication, Repair, and Recombination



Faithful copying is essential to the storage of genetic information. With the precision of a diligent monk copying an illuminated manuscript, a DNA polymerase (above) copies DNA strands, preserving the precise sequence of bases with very few errors. [(Left) The Pierpont Morgan Library/Art Resource.]

OUTLINE

28.1 DNA Replication Proceeds by the Polymerization of Deoxyribonucleoside Triphosphates Along a Template

28.2 DNA Unwinding and Supercoiling Are Controlled by Topoisomerases

28.3 DNA Replication Is Highly Coordinated

28.4 Many Types of DNA Damage Can Be Repaired

28.5 DNA Recombination Plays Important Roles in Replication, Repair, and Other Processes

Perhaps the most exciting aspect of the structure of DNA deduced by Watson and Crick was, as expressed in their words, that the "specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." A double helix separated into two single strands can be replicated because each strand serves as a template on which its complementary strand can be assembled (Figure 28.1). To preserve the information encoded in DNA through many cell divisions, copying of the genetic information must be extremely faithful. To replicate the human genome without mistakes, an error rate of less than 1 bp per 6×10^9 bp must be achieved. Such remarkable accuracy is achieved through a multilayered system of accurate DNA synthesis (which has an error rate of 1 per 10^3–10^4 bases inserted), proofreading during DNA synthesis (which reduces that error rate to approximately 1 per 10^6–10^7 bp), and postreplication mismatch repair (which reduces the error rate to approximately 1 per 10^9–10^{10} bp).

Even after DNA has been initially replicated, the genome is still not safe. Although DNA is remarkably robust, ultraviolet light as well as a range of

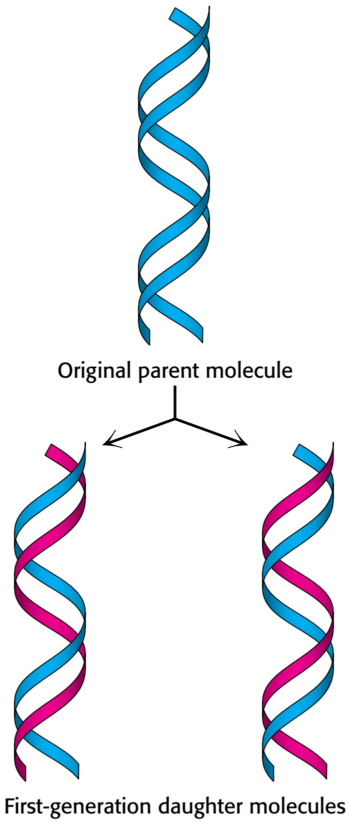


FIGURE 28.1 DNA replication. Each strand of one double helix (shown in blue) acts as a template for the synthesis of a new complementary strand (shown in red).

chemical species can damage DNA, introducing changes in the DNA sequence (mutations) or lesions that can block further DNA replication (Figure 28.2). All organisms contain DNA-repair systems that detect DNA damage and act to preserve the original sequence. Mutations in genes that encode components of DNA-repair systems are key factors in the development of cancer. Among the most potentially devastating types of DNA damage are double-stranded breaks in DNA. With both strands of the double helix broken in a local region, neither strand is intact to act as a template for future DNA synthesis. A mechanism used to repair such lesions relies on DNA recombination—that is, the reassortment of DNA sequences present on two different double helices. In addition to its role in DNA repair, recombination is crucial for the generation of genetic diversity in meiosis. Recombination is also the key to generating a highly diverse repertoire of genes for key molecules in the immune system (Chapter 34).

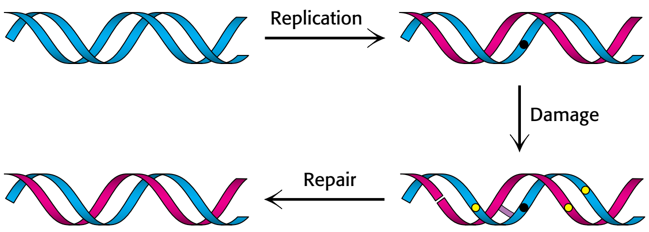


FIGURE 28.2 DNA Replication, damage, and repair. Some errors (shown as a black dot) may arise in the replication processes. Additional defects (shown in yellow), including modified bases, cross-links, and single- and double-strand breaks, are introduced into DNA by subsequent DNA-damaging reactions. Many of the errors are detected and subsequently repaired.

The bacterium *Deinococcus radiodurans* illustrates the extraordinary power of DNA repair systems. This bacterium was discovered in 1956 when scientists were studying the use of high doses of gamma radiation to sterilize canned meat. In some cases, the meat still spoiled due to the growth of a bacterial species that withstood doses of gamma radiation more than 1000 times larger than those that would kill a human being. Each *D. radiodurans* cell contains between 4 and 10 copies of its genome. Even when these bacterial chromosomes are broken into many fragments by the ionizing radiation, they can reassemble and recombine to regenerate the intact genome with essentially no loss of information (Figure 28.3). These cells can also survive extreme desiccation, that is, drying out, much better than other organisms. This ability is believed to be the selective advantage that favored the evolution of this and related species.

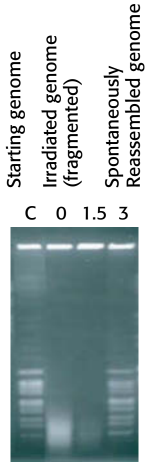

Starting genome
Irradiated genome (fragmented)
Spontaneously Reassembled genome
C
0
1.5
3

FIGURE 28.3 DNA repair in action. The bacterium *Deinococcus radiodurans* can reassemble its genome over the course of 3 hours after irradiation with gamma rays to fragment the genome into many pieces. To aid analysis, the genomic DNA samples were digested with a restriction enzyme that cuts at only a few sites within the genome. [Adapted from K. Zahradka et al., *Nature* 445:569–573, 2006.]

28.1 DNA Replication Proceeds by the Polymerization of Deoxyribonucleoside Triphosphates Along a Template

The base sequences of newly synthesized DNA must faithfully match the sequences of parent DNA. To achieve faithful replication, each strand within the parent double helix acts as a *template* for the synthesis of a new DNA strand with a complementary sequence. The building blocks for the synthesis of the new strands are deoxyribonucleoside triphosphates. They are added, one at a time, to the 3′ end of an existing strand of DNA.

Although this reaction is in principle quite simple, it is significantly complicated by specific features of the DNA double helix. First, the two strands of the double helix run in opposite directions. Because DNA strand synthesis always proceeds in the 5′-to-3′ direction, the DNA replication process must have special mechanisms to accommodate the oppositely directed strands. Second, the two strands of the double helix interact with one another in such a way that the bases, key templates for replication, are on the inside of the helix. Thus, the two strands must be separated from each other so as to generate appropriate templates. Finally, the two strands of the double helix wrap around each other. Thus, strand separation also entails the unwinding of the double helix. This unwinding creates supercoils that must themselves be resolved as replication continues. We begin with a consideration of the chemistry that underlies the formation of the phosphodiester backbone of newly synthesized DNA.

DNA polymerases require a template and a primer

DNA polymerases catalyze the formation of polynucleotide chains. Each incoming nucleoside triphosphate first forms an appropriate base pair with a base in the template. Only then does the DNA polymerase link the incoming base with the predecessor in the chain. Thus, *DNA polymerases are template-directed enzymes.*

DNA polymerases add nucleotides to the 3′ end of a polynucleotide chain. The polymerase catalyzes the nucleophilic attack by the 3′-hydroxyl-group terminus of the polynucleotide chain on the α phosphoryl group of the nucleoside triphosphate to be added (Figure 4.25). To initiate this reaction, DNA polymerases require a *primer* with a free 3′-hydroxyl group already base-paired to the template. They cannot start from scratch by adding nucleotides to a free single-stranded DNA template. RNA polymerase, in contrast, can initiate RNA synthesis without a primer, as we shall see in Chapter 29.

Primer
The initial segment of a polymer that is to be extended on which elongation depends.

Template
A sequence of DNA or RNA that directs the synthesis of a complementary sequence.

All DNA polymerases have structural features in common

The three-dimensional structures of a number of DNA polymerase enzymes are known. The first such structure was elucidated by Tom Steitz and coworkers, who determined the structure of the so-called *Klenow fragment* of DNA polymerase I from *E. coli* (Figure 28.4). This fragment comprises two main parts of the full enzyme, including the polymerase unit. This unit approximates the shape of a right hand with domains that are referred to as the fingers, the thumb, and the palm. In addition to the polymerase, the Klenow fragment includes a domain with 3′ → 5′ *exonuclease* activity that participates in proofreading and correcting the polynucleotide product.

DNA polymerases are remarkably similar in overall shape, although they differ substantially in detail. At least five structural classes have been identified; some of them are clearly homologous, whereas others appear to be the products of convergent evolution. In all cases, the finger and thumb domains wrap around DNA and hold it across the enzyme's active site, which comprises residues primarily from the palm domain. Furthermore, all DNA polymerases use similar strategies to catalyze the polymerase reaction, making use of a mechanism in which two metal ions take part.

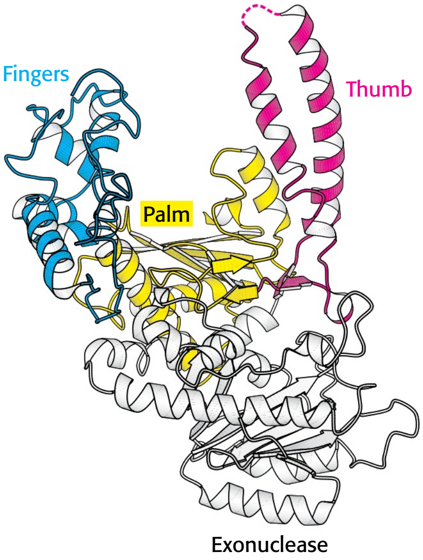


FIGURE 28.4 DNA polymerase structure. The first DNA polymerase structure determined was that of a fragment of *E. coli* DNA polymerase I called the Klenow fragment. *Notice* that, like other DNA polymerases, the polymerase unit resembles a right hand with fingers (blue), palm (yellow), and thumb (red). The Klenow fragment also includes an exonuclease domain that removes incorrect nucleotide bases. [Drawn from 1DPI.pdb.]

Two bound metal ions participate in the polymerase reaction

Like all enzymes with nucleoside triphosphate substrates, DNA polymerases require metal ions for activity. Examination of the structures of DNA polymerases with bound substrates and substrate analogs reveals the presence of two metal ions in the active site. One metal ion binds both the deoxynucleoside triphosphate (dNTP) and the 3′-hydroxyl group of the

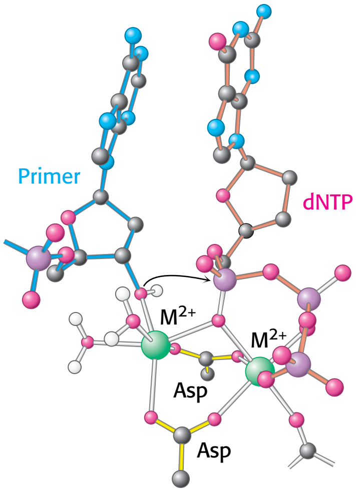


FIGURE 28.5 DNA polymerase mechanism. Two metal ions (typically, Mg^{2+}) participate in the DNA polymerase reaction. One metal ion coordinates the 3′-hydroxyl group of the primer, whereas the other metal ion interacts only with the dNTP. The phosphoryl group of the nucleoside triphosphate bridges between the two metal ions. The hydroxyl group of the primer attacks the phosphoryl group to form a new O—P bond.

primer, whereas the other interacts only with the dNTP (Figure 28.5). The two metal ions are bridged by the carboxylate groups of two aspartate residues in the palm domain of the polymerase. These side chains hold the metal ions in the proper positions and orientations. The metal ion bound to the primer activates the 3′-hydroxyl group of the primer, facilitating its attack on the α phosphoryl group of the dNTP substrate in the active site. The two metal ions together help stabilize the negative charge that accumulates on the pentacoordinate transition state. The metal ion initially bound to dNTP stabilizes the negative charge on the pyrophosphate product.

The specificity of replication is dictated by complementarity of shape between bases

DNA must be replicated with high fidelity. Each base added to the growing chain should, with high probability, be the Watson–Crick complement of the base in the corresponding position in the template strand. The binding of the dNTP containing the proper base is favored by the formation of a base pair with its partner on the template strand. Although hydrogen bonding contributes to the formation of this base pair, overall shape complementarity is crucial. Studies show that a nucleotide with a base that is very similar in shape to adenine but lacks the ability to form base-pairing hydrogen bonds can still direct the incorporation of thymidine, both in vitro and in vivo (Figure 28.6).

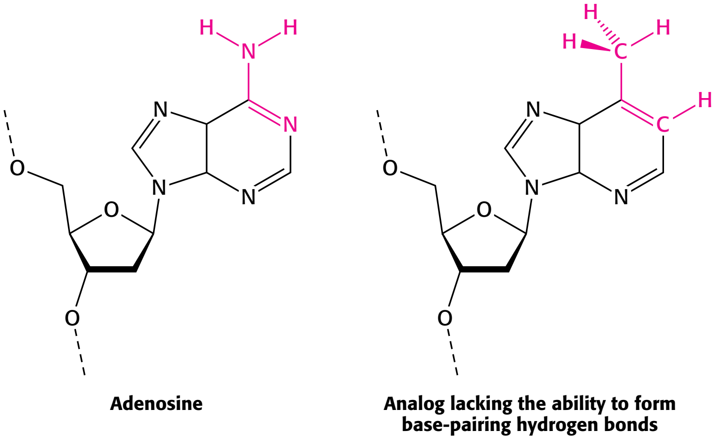

Adenosine
Analog lacking the ability to form base-pairing hydrogen bonds

FIGURE 28.6 Shape complementarity. The base analog on the right has the same shape as adenosine, but groups that form hydrogen bonds between base pairs have been replaced by groups (shown in red) not capable of hydrogen bonding. Nonetheless, studies reveal that, when incorporated into the template strand, this analog directs the insertion of thymidine in DNA replication.

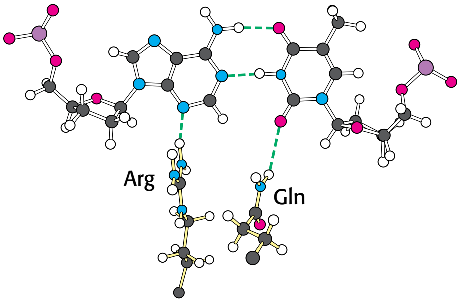


FIGURE 28.7 Minor-groove interactions. DNA polymerases donate two hydrogen bonds to base pairs in the minor groove. Hydrogen-bond acceptors are present in these two positions for all Watson–Crick base pairs, including the A—T base pair shown.

An examination of the crystal structures of various DNA polymerases reveals why shape complementarity is so important. First, residues of the enzyme form hydrogen bonds with *the minor-groove side of the base pair in the active site* (Figure 28.7). In the minor groove, hydrogen-bond acceptors are present in the same positions for all Watson–Crick base pairs. These interactions act as a "ruler" that measures whether a properly spaced base pair has formed in the active site.

Second, DNA polymerases close down around the incoming dNTP (Figure 28.8). The binding of a deoxyribonucleoside triphosphate into the active site of a DNA polymerase triggers a conformational change: the finger domain rotates to form a tight pocket into which only a properly shaped base pair will readily fit. Many of the residues lining this pocket are important to ensure the efficiency and fidelity of DNA synthesis. For example, mutation of a conserved tyrosine residue that forms part of the pocket results in a polymerase that is approximately 40 times as error prone as the parent polymerase.

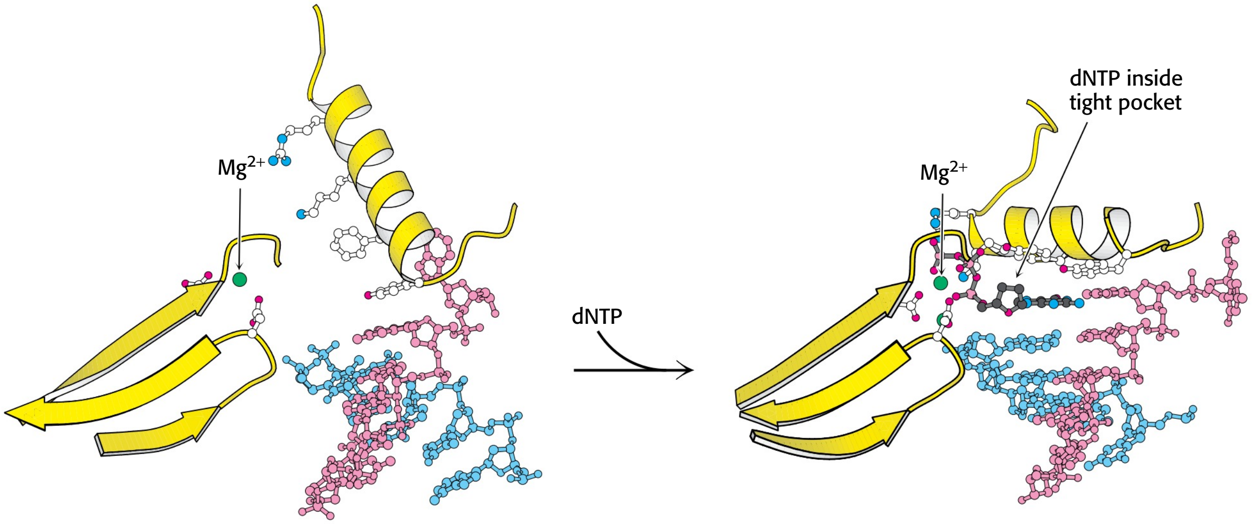


FIGURE 28.8 Shape selectivity. The binding of a deoxyribonucleoside triphosphate (dNTP) to DNA polymerase induces a conformational change, generating a tight pocket for the base pair consisting of the dNTP and its partner on the template strand. Such a conformational change is possible only when the dNTP corresponds to the Watson–Crick partner of the template base. [Drawn from 2BDP.pdb and 1T7P.pdb.]

An RNA primer synthesized by primase enables DNA synthesis to begin

DNA polymerases cannot initiate DNA synthesis without a primer, a section of nucleic acid having a free 3′ end that forms a double helix with the template. How is this primer formed? An important clue came from the observation that RNA synthesis is essential for the initiation of DNA synthesis. In fact, *RNA primes the synthesis of DNA*. An RNA polymerase called *primase* synthesizes a short stretch of RNA (about five nucleotides) that is complementary to one of the template DNA strands (Figure 28.9). Primase, like other RNA polymerases, can initiate synthesis without a primer. After DNA synthesis has been initiated, the short stretch of RNA is removed by hydrolysis and replaced by DNA.

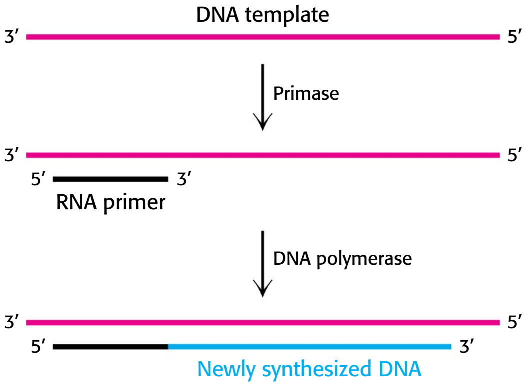


FIGURE 28.9 Priming. DNA replication is primed by a short stretch of RNA that is synthesized by primase, an RNA polymerase. The RNA primer is removed at a later stage of replication.

One strand of DNA is made continuously, whereas the other strand is synthesized in fragments

Both strands of parental DNA serve as templates for the synthesis of new DNA. The site of DNA synthesis is called the *replication fork* because the complex formed by the newly synthesized daughter helices arising from the parental duplex resembles a two-pronged fork. Recall that the two strands are antiparallel; that is, they run in opposite directions. During DNA replication, both daughter strands appear on cursory examination to grow in the same direction. However, all known DNA polymerases synthesize DNA in the 5′ → 3′ direction but not in the 3′ → 5′ direction. How then does one of the daughter DNA strands appear to grow in the 3′ → 5′ direction?

This dilemma was resolved by Reiji Okazaki, who found that *a significant proportion of newly synthesized DNA exists as small fragments*. These units of about a thousand nucleotides (called *Okazaki fragments*) are present briefly in the vicinity of the replication fork (Figure 28.10).

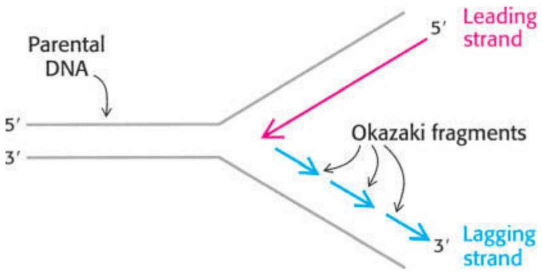


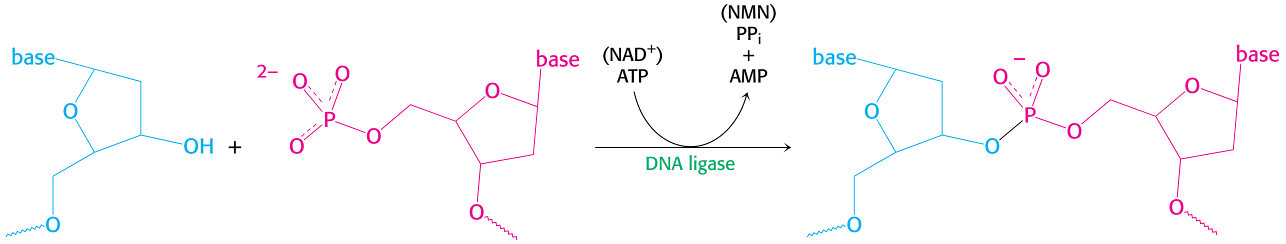
FIGURE 28.10 Okazaki fragments. At a replication fork, both strands are synthesized in the 5′ → 3′ direction. The leading strand is synthesized continuously, whereas the lagging strand is synthesized in short pieces termed Okazaki fragments.

As replication proceeds, these fragments become covalently joined through the action of the enzyme DNA ligase to form a continuous daughter strand. The other new strand is synthesized continuously. The strand formed from Okazaki fragments is termed the *lagging strand,* whereas the one synthesized without interruption is the *leading strand.* The discontinuous assembly of the lagging strand enables $5' \rightarrow 3'$ polymerization at the nucleotide level to give rise to overall growth in the $3' \rightarrow 5'$ direction.

DNA ligase joins ends of DNA in duplex regions

The joining of Okazaki fragments requires an enzyme that catalyzes the joining of the ends of two DNA chains. The existence of circular DNA molecules also points to the existence of such an enzyme. In 1967, scientists in several laboratories simultaneously discovered *DNA ligase. This enzyme catalyzes the formation of a phosphodiester linkage between the 3′-hydroxyl group at the end of one DNA chain and the 5′-phosphoryl group at the end of the other* (Figure 28.11). An energy source is required to drive this thermodynamically uphill reaction. In eukaryotes and archaea, ATP is the energy source. In bacteria, NAD^+ typically plays this role.

FIGURE 28.11 DNA ligase reaction. DNA ligase catalyzes the joining of one DNA strand with a free 3′-hydroxyl group to another with a free 5′-phosphoryl group. In eukaryotes and archaea, ATP is cleaved to AMP and PP_i to drive this reaction. In bacteria, NAD^+ is cleaved to AMP and nicotinamide mononucleotide (NMN).



DNA ligase cannot link two molecules of single-stranded DNA or circularize single-stranded DNA. Rather, *ligase seals breaks in double-stranded DNA molecules.* The enzyme from *E. coli* ordinarily forms a phosphodiester bridge only if there are at least a few bases of single-stranded DNA on the end of a double-stranded fragment that can come together with those on another fragment to form base pairs. Ligase encoded by T4 bacteriophage can link two blunt-ended double-helical fragments, a capability that is exploited in recombinant DNA technology.

The separation of DNA strands requires specific helicases and ATP hydrolysis

For a double-stranded DNA molecule to replicate, the two strands of the double helix must be separated from each other, at least locally. This separation allows each strand to act as a template on which a new polynucleotide chain can be assembled. Specific enzymes, termed *helicases,* utilize the energy of ATP hydrolysis to power strand separation.

Helicases are a large and diverse family of enzymes taking part in many biological processes. The helicases in DNA replication are typically oligomers containing six subunits that form a ring structure. The structure of one such helicase, that from bacteriophage T7, has been a source of considerable insight into the helicase mechanism (Figure 28.12). Each of the subunits within this hexameric structure has a core structure that includes a P-loop NTPase domain (Figure 9.49). In addition to the P-loop, each subunit has two loops that extend toward the center of the ring structure and interact with DNA. Each subunit interacts closely with its two neighbors within the ring structure. Closer examination of this structure reveals that the ring deviates significantly from six-fold symmetry. This deviation is even more

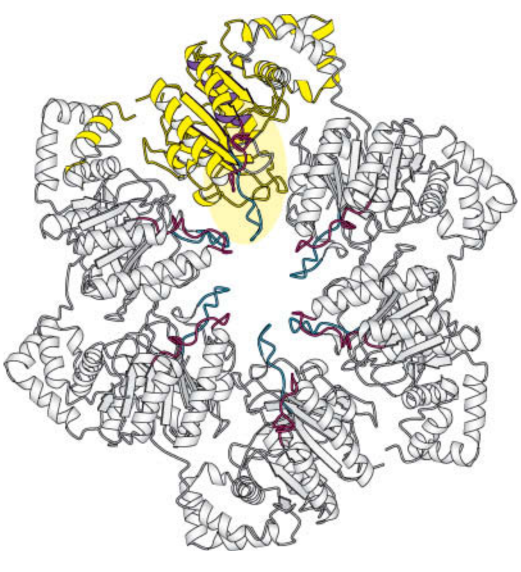
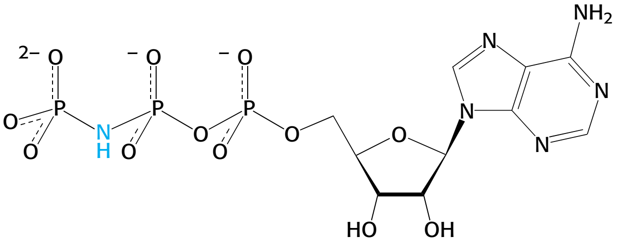


FIGURE 28.12 Helicase structure. The structure of the hexameric helicase from bacteriophage T7. One of the six subunits is shown in yellow with the P-loop NTPase shown in purple. The loops that participate in DNA binding are highlighted by a yellow oval. *Notice* that each subunit interacts closely with its neighbors and that the DNA-binding loops line the hole in the center of the structure. [Drawn from 1E0K.pdb.]

apparent when helicase is crystallized in the presence of the nonhydrolyzable ATP analog AMP-PNP.



AMP-PNP

The AMP-PNP binds to only four of the six subunits within the ring (Figure 28.13). Furthermore, the four nucleotide-binding sites are not identical but fall into two classes. One class appears to be well positioned to bind ATP but not catalyze its hydrolysis, whereas the other class is more well suited to catalyze the hydrolysis but not release the hydrolysis products. The classes are analogous to myosin's two different conformations—one for binding ATP and one for hydrolyzing it (Section 9.4). Finally, the six subunits fall into three classes with regard to their orientation with respect to the overall ring structure, with differences in rotation around an axis in the plane of the ring of approximately 30°. These differences in orientation affect the position of the two DNA-binding loops in each subunit.

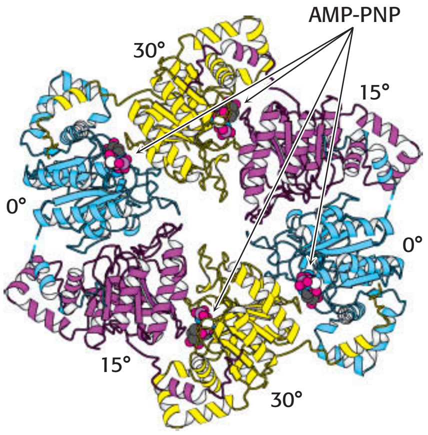


FIGURE 28.13 Helicase asymmetry. The structure of the T7 helicase complexes with the ATP analog AMP-PNP is shown. The three classes of helicase subunits are shown in blue, red, and yellow. The rotation relative to the plane of the hexamer is shown for each subunit. *Notice* that only four of the subunits, those shown in blue and yellow, bind AMP-PNP. [Drawn from 1E0K.pdb.]

These observations are consistent with the following mechanism for the helicase (Figure 28.14). Only a single strand of DNA can fit through the center on the ring. This single strand binds to loops on two adjacent subunits, one of which has bound ATP and the other of which has bound ADP + P_i. The binding of ATP to the domains that initially had no bound nucleotides leads to a conformational change within the entire hexamer, leading to the release of ADP + P_i from two subunits and the binding of the single-stranded DNA by one of the domains that just bound ATP. This conformational change pulls the DNA through the center of the hexamer. The protein acts as a wedge, forcing the two strands of the double helix apart. This cycle then repeats itself, moving two bases along the DNA strand with each cycle.

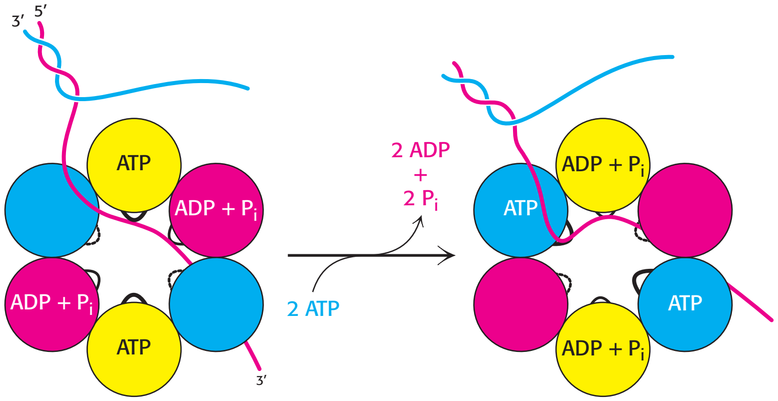
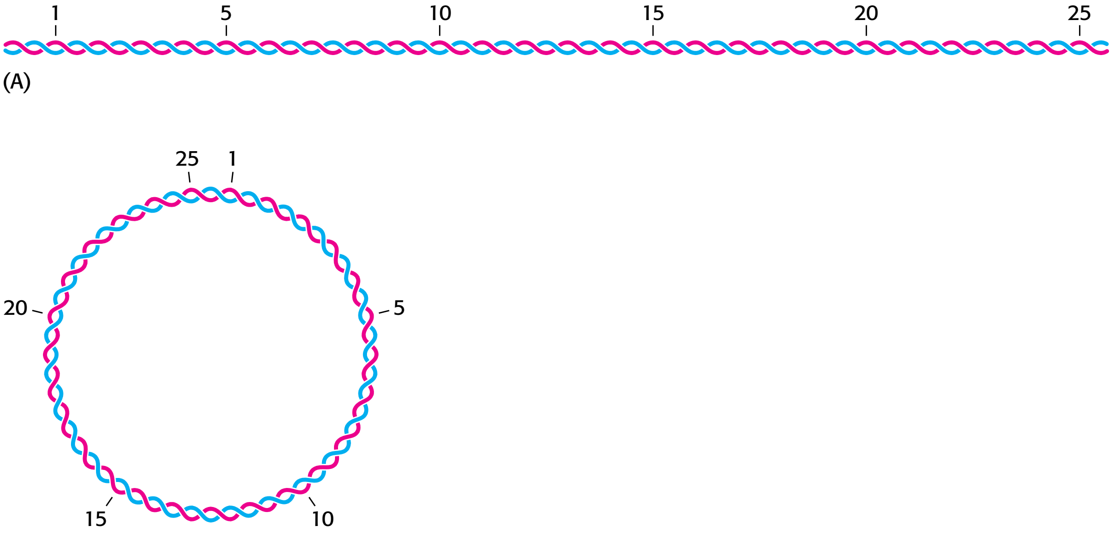


FIGURE 28.14 Helicase mechanism. One of the strands of the double helix passes through the hole in the center of the helicase, bound to the loops of two adjacent subunits. Two of the subunits do not contain bound nucleotides. On the binding of ATP to these two subunits and the release of ADP + P_i from two other subunits, the helicase hexamer undergoes a conformational change, pulling the DNA through the helicase. The helicase acts as a wedge to force separation of the two strands of DNA.

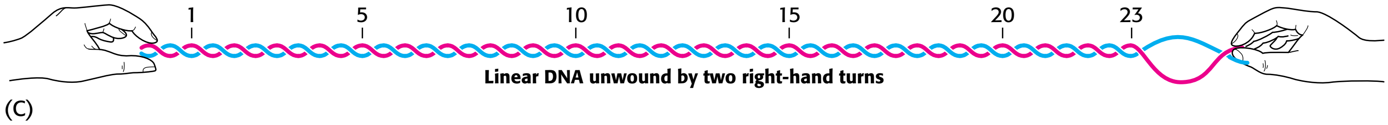
28.2 DNA Unwinding and Supercoiling Are Controlled by Topoisomerases

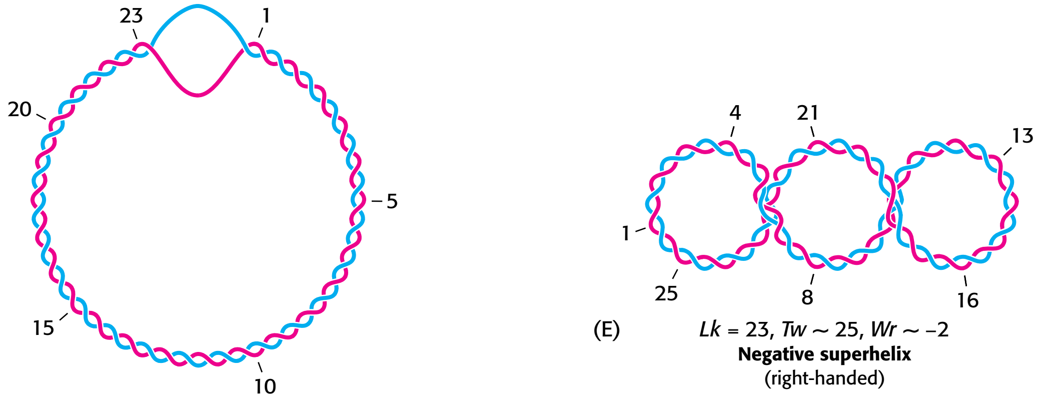
As a helicase moves along unwinding DNA, the DNA in front of the helicase will become overwound in the absence of other changes. As discussed in Section 4.2, DNA double helices that are torsionally stressed tend to fold up on themselves to form tertiary structures created by *supercoiling*. We will first consider the supercoiling of DNA in quantitative terms and then turn to topoisomerases, enzymes that can directly modulate DNA winding and supercoiling. Supercoiling is most readily understood by considering circular DNA molecules, but it also applies to linear DNA molecules constrained to be in loops by other means. Most DNA molecules inside cells are subject to supercoiling.

Consider a linear 260-bp DNA duplex in the B-DNA form (Figure 28.15A). Because the number of base pairs per turn in an unstressed DNA molecule averages 10.4, this linear DNA molecule has 25 (260/10.4) turns. The ends of this helix can be joined to produce a *relaxed* circular DNA (Figure 28.15B). A different circular DNA can be formed by unwinding



(B) $Lk = 25$, $Tw = 25$, $Wr = 0$
Relaxed circle





(D) $Lk = 23$, $Tw \sim 23$, $Wr \sim 0$
Unwound circle

(E) $Lk = 23$, $Tw \sim 25$, $Wr \sim -2$
Negative superhelix
(right-handed)

FIGURE 28.15 Linking number. The relations between the linking number *(Lk)*, twisting number *(Tw)*, and writhing number *(Wr)* of a circular DNA molecule revealed schematically. [Information from W. Saenger, *Principles of Nucleic Acid Structure* (Springer Verlag, 1984), p. 452.]

the linear duplex by two turns before joining its ends (Figure 28.15C). What is the structural consequence of unwinding before ligation? Two limiting conformations are possible. The DNA can fold into a structure containing 23 turns of B helix and an unwound loop (Figure 28.15D). Alternatively, the double helix can fold up to cross itself. Such crossings are called *supercoils*. In particular, a supercoiled structure with 25 turns of B helix and 2 turns of *right-handed* (termed *negative*) superhelix can be formed (Figure 28.15E).

Supercoiling markedly alters the overall form of DNA. *A supercoiled DNA molecule is more compact than a relaxed DNA molecule of the same length.* Hence, supercoiled DNA moves faster than relaxed DNA when analyzed by centrifugation or electrophoresis. Unwinding will cause supercoiling in circular DNA molecules, whether covalently closed or constrained in closed configurations by other means.

The linking number of DNA, a topological property, determines the degree of supercoiling

Our understanding of the conformation of DNA is enriched by concepts drawn from topology, a branch of mathematics dealing with structural properties that are unchanged by deformations such as stretching and bending. A key topological property of a circular DNA molecule is its *linking number (Lk)*, which is equal to the number of times that a strand of DNA winds in the right-handed direction around the helix axis when the axis lies in a plane, as in Figure 28.15A. For the relaxed DNA shown in Figure 28.15B, $Lk = 25$. For the partly unwound molecule shown in part D and the supercoiled one shown in part E, $Lk = 23$ because the linear duplex was unwound two complete turns before closure. Molecules differing only in linking number are *topological isomers,* or *topoisomers,* of one another. *Topoisomers of DNA can be interconverted only by cutting one or both DNA strands and then rejoining them.*

The unwound DNA and supercoiled DNA shown in Figure 28.15D and E are topologically identical but geometrically different. They have the same value of Lk but differ in *twist (Tw)* and *writhe (Wr)*. Although the rigorous definitions of twist and writhe are complex, twist is a measure of the helical winding of the DNA strands around each other, whereas writhe is a measure of the coiling of the axis of the double helix—that is, supercoiling. A right-handed coil is assigned a negative number (negative supercoiling) and a left-handed coil is assigned a positive number (positive supercoiling).

Is there a relation between Tw and Wr? Indeed, there is. Topology tells us that the sum of Tw and Wr is equal to Lk.

$$Lk = Tw + Wr$$

In Figure 28.15, the partly unwound circular DNA has $Tw \sim 23$, meaning the helix has 23 turns and $Wr \sim 0$, meaning the helix has not crossed itself to create a supercoil. The supercoiled DNA, however has $Tw \sim 25$ and $Wr \sim -2$. These forms can be interconverted without cleaving the DNA chain because they have the same value of Lk—namely, 23. The partitioning of Lk (which must be an integer) between Tw and Wr (which need not be integers) is determined by energetics. The free energy is minimized when about 70% of the change in Lk is expressed in Wr and 30% is expressed in Tw. Hence, the most stable form would be one with $Tw = 24.4$ and $Wr = -1.4$. Thus, *a lowering of* Lk *causes both right-handed (negative) supercoiling of the DNA axis and unwinding of the duplex.* Topoisomers differing by just 1 in Lk, and consequently

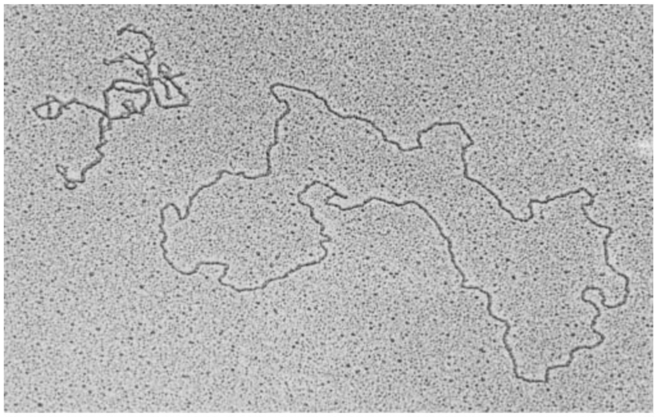


FIGURE 28.16 Topoisomers. An electron micrograph showing negatively supercoiled and relaxed DNA. [Courtesy of Dr. Jack Griffith.]

by 0.7 in *Wr,* can be readily separated by agarose gel electrophoresis because their hydrodynamic volumes are quite different; *supercoiling condenses DNA* (Figure 28.16).

Topoisomerases prepare the double helix for unwinding

Most naturally occurring DNA molecules are negatively supercoiled. What is the basis for this prevalence? As already stated, negative supercoiling arises from the unwinding or underwinding of the DNA. In essence, negative supercoiling prepares DNA for processes requiring separation of the DNA strands, such as replication. Positive supercoiling condenses DNA as effectively, but it makes strand separation more difficult.

The presence of supercoils in the immediate area of unwinding would, however, make unwinding difficult. Therefore, negative supercoils must be continuously removed, and the DNA relaxed, as the double helix unwinds. Specific enzymes called *topoisomerases* that introduce or eliminate supercoils were discovered by James Wang and Martin Gellert. *Type I topoisomerases* catalyze the relaxation of supercoiled DNA, a thermodynamically favorable process. *Type II topoisomerases* utilize free energy from ATP hydrolysis to add negative supercoils to DNA. Both type I and type II topoisomerases play important roles in DNA replication as well as in transcription and recombination.

These enzymes alter the linking number of DNA by catalyzing a three-step process: (1) the *cleavage* of one or both strands of DNA, (2) the *passage* of a segment of DNA through this break, and (3) the *resealing* of the DNA break. Type I topoisomerases cleave just one strand of DNA, whereas type II enzymes cleave both strands. The two types of enzymes have several common features, including the use of key tyrosine residues to form covalent links to the polynucleotide backbone that is transiently broken.

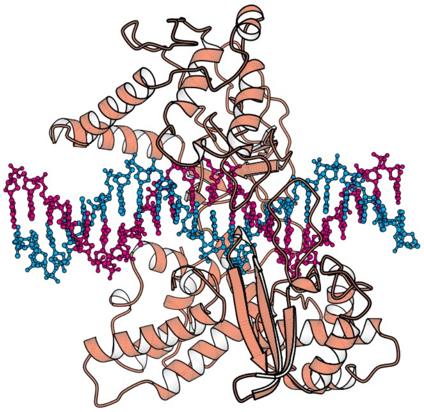


FIGURE 28.17 Structure of topoisomerase I. The structure of a complex between a fragment of human topoisomerase I and DNA is shown. *Notice* that DNA lies in a central cavity within the enzyme. [Drawn from 1EJ9.pdb.]

Type I topoisomerases relax supercoiled structures

The three-dimensional structures of several type I topoisomerases have been determined (Figure 28.17). These structures reveal many features of the reaction mechanism. Human type I topoisomerase comprises four domains, which are arranged around a central cavity having a diameter of 20 Å, just the correct size to accommodate a double-stranded DNA molecule. This cavity also includes a tyrosine residue (Tyr 723), which acts as a nucleophile to cleave the DNA backbone in the course of catalysis.

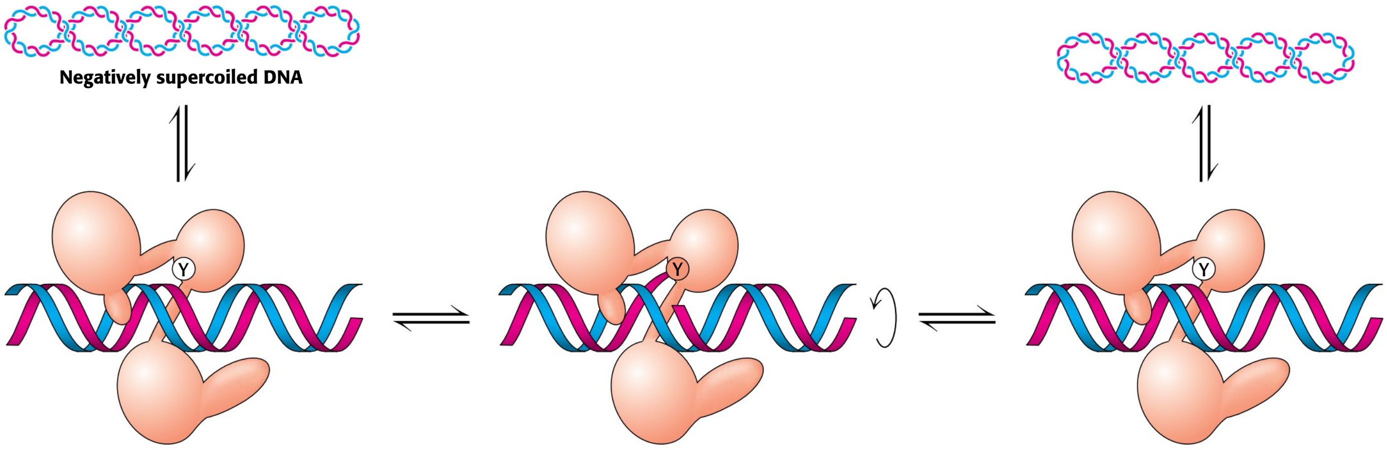
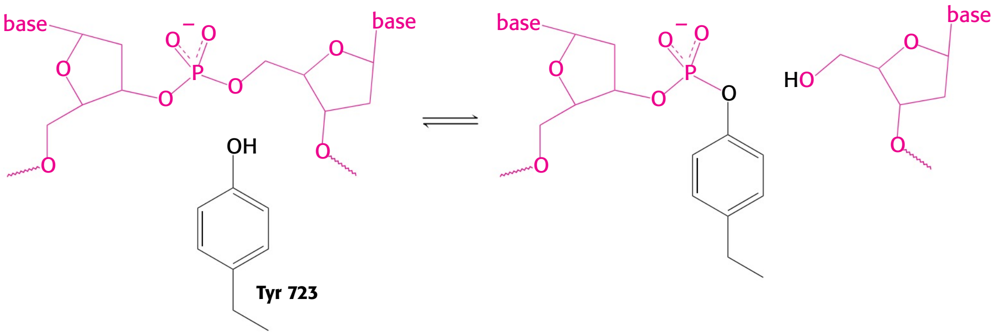


FIGURE 28.18 Topoisomerase I mechanism. On binding to DNA, topoisomerase I cleaves one strand of the DNA by means of a tyrosine (Y) residue attacking a phosphoryl group. When the strand has been cleaved, it rotates in a controlled manner around the other strand. The reaction is completed by religation of the cleaved strand. This process results in partial or complete relaxation of a supercoiled plasmid.

From analyses of these structures and the results of other studies, the relaxation of negatively supercoiled DNA molecules is known to proceed in the following manner (Figure 28.18). First, the DNA molecule binds inside the cavity of the topoisomerase. The hydroxyl group of tyrosine 723 attacks a phosphoryl group on one strand of the DNA backbone to form a phosphodiester linkage between the enzyme and the DNA, cleaving the DNA and releasing a free 5′-hydroxyl group.



With the backbone of one strand cleaved, the DNA can now rotate around the remaining strand, its movement driven by the release of energy stored because of the supercoiling. The rotation of the DNA unwinds the supercoils. The enzyme controls the rotation so that the unwinding is not rapid. The free hydroxyl group of the DNA attacks the phosphotyrosine residue to reseal the backbone and release tyrosine. The DNA is then free to dissociate from the enzyme. Thus, reversible cleavage of one strand of supercoiled DNA allows controlled rotation to partly relax the supercoils.

Type II topoisomerases can introduce negative supercoils through coupling to ATP hydrolysis

Supercoiling requires an input of energy because a supercoiled molecule, in contrast with its relaxed counterpart, is torsionally stressed. The introduction of an additional supercoil into a 3000-bp plasmid typically requires about 30 kJ mol^{-1} (7 kcal mol^{-1}).

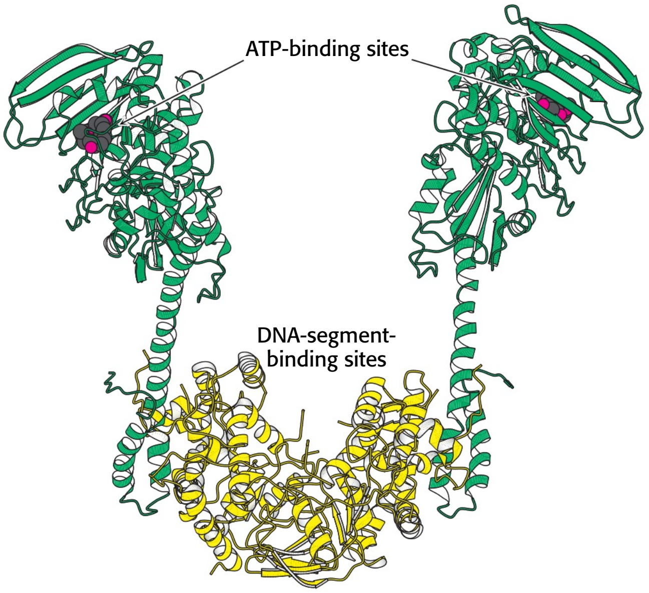


FIGURE 28.19 Structure of topoisomerase II. The dimeric structure of a typical topoisomerase II, that from the archaeon *Sulfolobus shibatae. Notice* that each half of the enzyme has one domain (shown in yellow) that contains a region for binding a DNA double helix and another domain (shown in green) that contains ATP binding sites. [Drawn from 2ZBK.pdb.]

Supercoiling can be catalyzed by type II topoisomerases. These elegant molecular machines couple the binding and hydrolysis of ATP to the directed passage of one DNA double helix through another, temporarily cleaved DNA double helix. These enzymes have several mechanistic features in common with the type I topoisomerases.

Topoisomerase II molecules are dimeric with a large internal cavity (Figure 28.19). The large cavity has gates at both the top and the bottom that are crucial to topoisomerase action. The reaction begins with the binding of one double helix (hereafter referred to as the G, for gate, segment) to the enzyme (Figure 28.20). Each strand is positioned next to a tyrosine residue, one from each monomer, capable of forming a covalent linkage with the DNA backbone. This complex then loosely binds a second DNA double helix (hereafter referred to as the T, for transported, segment). Each monomer of the enzyme has a domain that binds ATP; this ATP binding leads to a conformational change that strongly favors the coming together of the two domains. As these domains come closer together, they trap the bound T segment. This conformational change also forces the separation and cleavage of the two strands of the G segment. Each strand is linked to the enzyme by a tyrosine–phosphodiester bond. Unlike the type I enzymes,

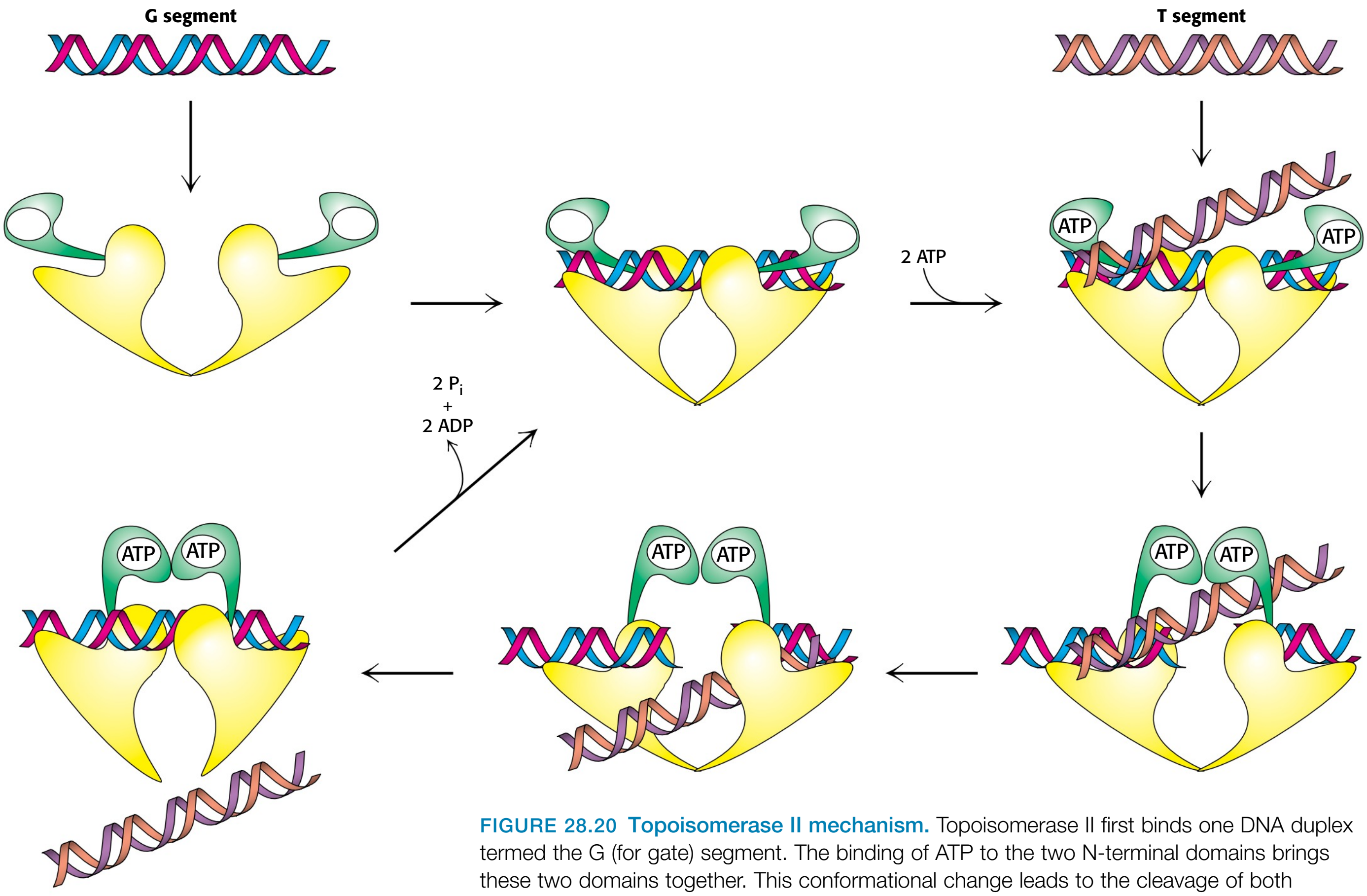
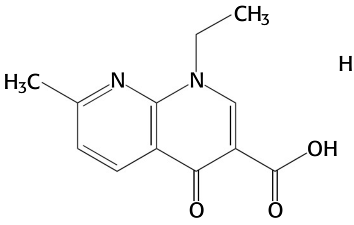


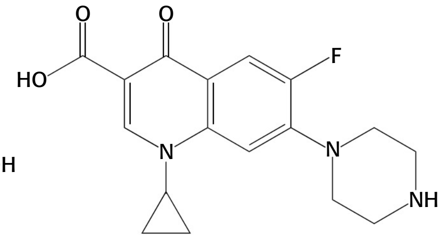
FIGURE 28.20 Topoisomerase II mechanism. Topoisomerase II first binds one DNA duplex termed the G (for gate) segment. The binding of ATP to the two N-terminal domains brings these two domains together. This conformational change leads to the cleavage of both strands of the G segment and the binding of an additional DNA duplex, the T segment. This T segment then moves through the break in the G segment and out the bottom of the enzyme. The hydrolysis of ATP resets the enzyme with the G segment still bound. The orientation of the entering T segment results in the introduction of negative supercoils.

the type II topoisomerases hold the DNA tightly so that it cannot rotate. The T segment then passes through the cleaved G segment and into the large central cavity. The ligation of the G segment leads to the release of the T segment through the gate at the bottom of the enzyme. The hydrolysis of ATP and the release of ADP and orthophosphate allow the ATP-binding domains to separate, preparing the enzyme to bind another T segment. The overall process leads to a decrease in the linking number by two.

The bacterial topoisomerase II (often called DNA gyrase) is the target of several antibiotics that inhibit the prokaryotic enzyme much more than the eukaryotic one. *Novobiocin* blocks the binding of ATP to gyrase. *Nalidixic acid* and *ciprofloxacin,* in contrast, interfere with the breakage and rejoining of DNA chains. These two gyrase inhibitors are widely used to treat urinary tract and other infections including those due to *Bacillus anthracis* (anthrax). *Camptothecin,* an antitumor agent, inhibits human topoisomerase I by stabilizing the form of the enzyme covalently linked to DNA.



Nalidixic acid



Ciprofloxacin

28.3 DNA Replication Is Highly Coordinated

DNA replication must be very rapid, given the sizes of the genomes and the rates of cell division. The *E. coli* genome contains 4.6 million base pairs and is copied in less than 40 minutes. Thus, 2000 bases are incorporated per second. Enzyme activities must be highly coordinated to replicate entire genomes precisely and rapidly.

We begin our consideration of the coordination of DNA replication by looking at *E. coli,* which has been extensively studied. For this organism with a relatively small genome, replication begins at a single site and continues around the circular chromosome. The coordination of eukaryotic DNA replication is much more complex because there are many initiation sites throughout the genome and an additional enzyme is needed to replicate the ends of linear chromosomes.

DNA replication requires highly processive polymerases

Replicative polymerases are characterized by their *very high catalytic potency, fidelity, and processivity. Processivity* refers to the ability of an enzyme to catalyze many consecutive reactions without releasing its substrate. These polymerases are assemblies of many subunits that have evolved to grasp their templates and not let go until many nucleotides have been added. The source of the processivity was revealed by the determination of the three-dimensional structure of the β_2 subunit of the *E. coli* replicative polymerase called DNA polymerase III (Figure 28.21). This unit keeps the polymerase associated with the DNA double helix. It has the form of a star-shaped ring. A 35-Å-diameter hole in its center can readily accommodate a duplex DNA molecule and leaves enough space between the DNA and the protein to allow rapid sliding during replication. To achieve a

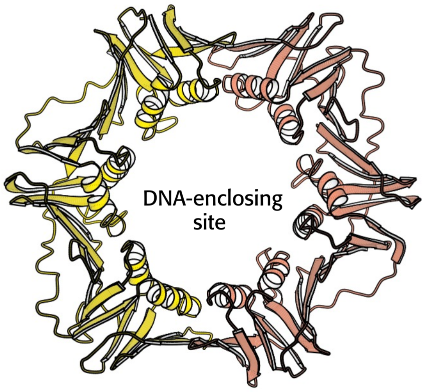


FIGURE 28.21 Structure of a sliding DNA clamp. The dimeric β subunit of DNA polymerase III forms a ring that surrounds the DNA duplex. *Notice* the central cavity through which the DNA template slides. Clasping the DNA molecule in the ring, the polymerase enzyme is able to move without falling off the DNA substrate. [Drawn from 2POL.pdb.]

catalytic rate of 1000 nucleotides polymerized per second requires that 100 turns of duplex DNA (a length of 3400 Å, or 0.34 mm) slide through the central hole of β_2 per second. Thus, β_2 plays a key role in replication by serving as a *sliding DNA clamp*.

How does DNA become entrapped inside the sliding clamp? Replicative polymerases also include assemblies of subunits that function as *clamp loaders*. These enzymes grasp the sliding clamp and, utilizing the energy of ATP binding, pull apart one of the interfaces between the two subunits of the sliding clamp. DNA can move through the gap, inserting itself through the central hole. ATP hydrolysis then releases the clamp, which closes around the DNA.

Processive enzyme

From the Latin *procedere,* "to go forward."

An enzyme that catalyzes multiple rounds of the elongation or digestion of a polymer while the polymer stays bound. A *distributive enzyme,* in contrast, releases its polymeric substrate between successive catalytic steps.

The leading and lagging strands are synthesized in a coordinated fashion

Replicative polymerases such as DNA polymerase III synthesize the leading and lagging strands simultaneously at the replication fork (Figure 28.22). DNA polymerase III begins the synthesis of the leading strand starting from the RNA primer formed by primase. The duplex DNA ahead of the polymerase is unwound by a hexameric helicase called DnaB. Copies of single-stranded-binding protein (SSB) bind to the unwound strands, keeping the strands separated so that both strands can serve as templates. The leading strand is synthesized continuously by polymerase III. Topoisomerase II concurrently introduces right-handed (negative) supercoils to avert a topological crisis.

The mode of synthesis of the lagging strand is necessarily more complex. As mentioned earlier, the lagging strand is synthesized in fragments so that $5' \rightarrow 3'$ polymerization leads to overall growth in the $3' \rightarrow 5'$ direction. Yet the synthesis of the lagging strand is coordinated with the synthesis of the leading strand. How is this coordination accomplished? Examination of the subunit composition of the DNA polymerase III holoenzyme reveals an elegant solution (Figure 28.23). The holoenzyme includes two copies of the polymerase core enzyme, which consists of the DNA polymerase itself (the α subunit); the ε subunit, a 3′-to-5′ proofreading exonuclease; another subunit called θ; and two copies of the dimeric β-subunit sliding clamp. The core enzymes are linked to a central structure having the subunit composition $\gamma\tau_2\delta\delta'\chi\phi$. The $\gamma\tau_2\delta\delta'$ complex is the clamp loader, and the χ and ϕ subunits

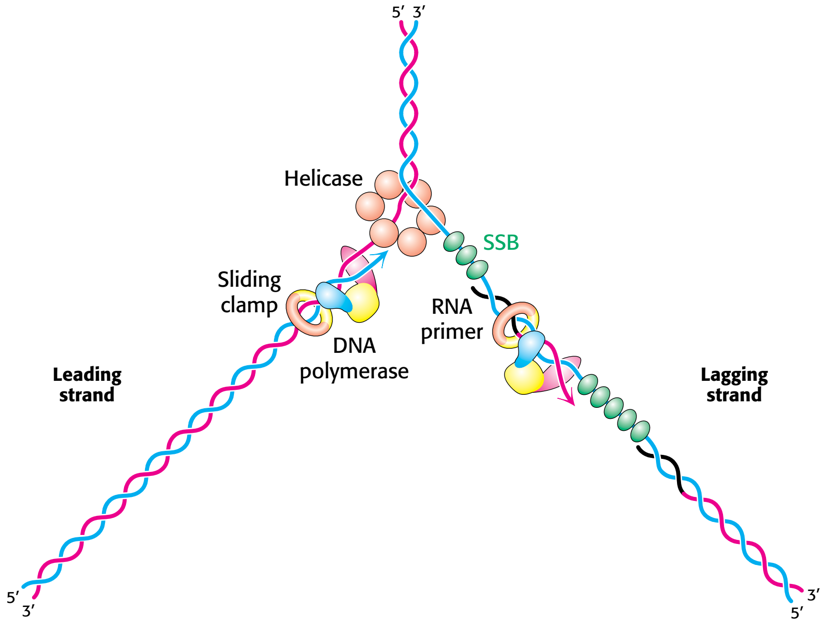


FIGURE 28.22 **Replication fork.** A schematic view of the arrangement of DNA polymerase III and associated enzymes and proteins present in the replication of DNA. The helicase separates the two strands of the parent double helix, allowing DNA polymerases to use each strand as a template for DNA synthesis. Abbreviation: SSB, single-stranded-binding protein.

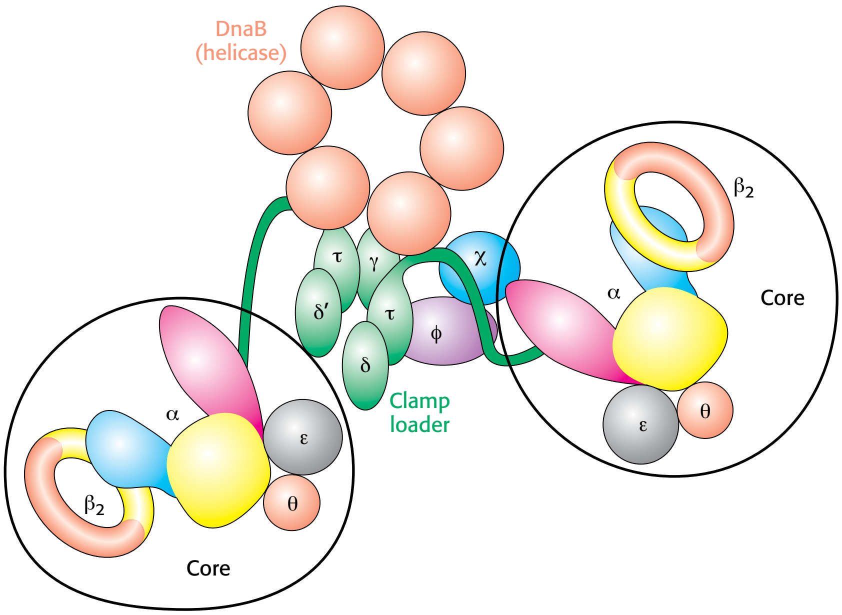


FIGURE 28.23 DNA polymerase holoenzyme. Each holoenzyme consists of two copies of the polymerase core enzyme, which comprises the α, ε, and θ subunits and two copies of the β subunit, linked to a central structure. The central structure includes the clamp-loader complex and the hexameric helicase DnaB.

interact with the single-stranded-DNA–binding protein. The entire apparatus interacts with the hexameric helicase DnaB. Eukaryotic replicative polymerases have similar, albeit slightly more complicated, subunit compositions and structures.

The lagging-strand template is looped out so that it passes through the polymerase site in one subunit of a dimeric polymerase III in the same direction as that of the leading-strand template in the other subunit, 5′ → 3′. DNA polymerase III lets go of the lagging-strand template after adding about 1000 nucleotides by releasing the sliding clamp. A new loop is then formed, a sliding clamp is added, and primase again synthesizes a short stretch of RNA primer to initiate the formation of another Okazaki fragment. This mode of replication has been termed the *trombone model* because the size of the loop lengthens and shortens like the slide on a trombone (Figure 28.24).

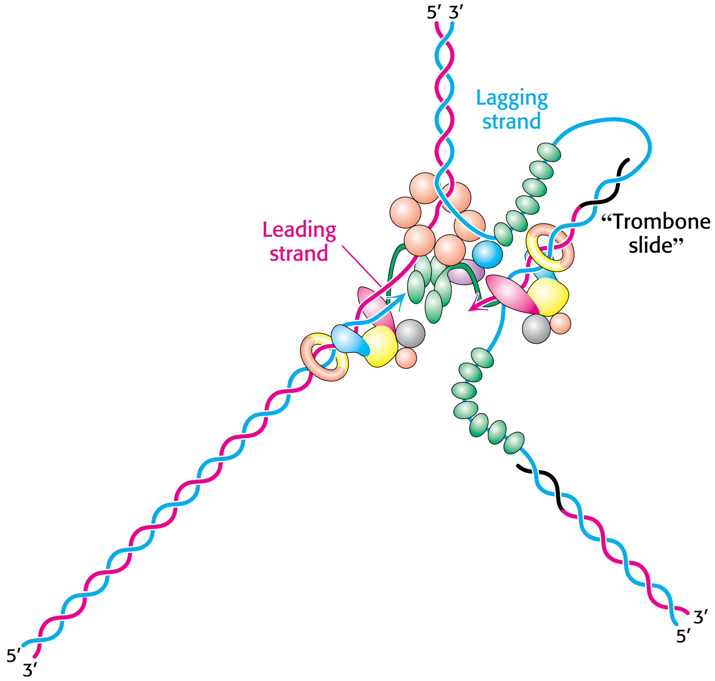


FIGURE 28.24 Trombone model. The replication of the leading and lagging strands is coordinated by the looping out of the lagging strand to form a structure that acts somewhat as a trombone slide, growing as the replication fork moves forward. When the polymerase on the lagging strand reaches a region that has been replicated, the sliding clamp is released and a new loop is formed.

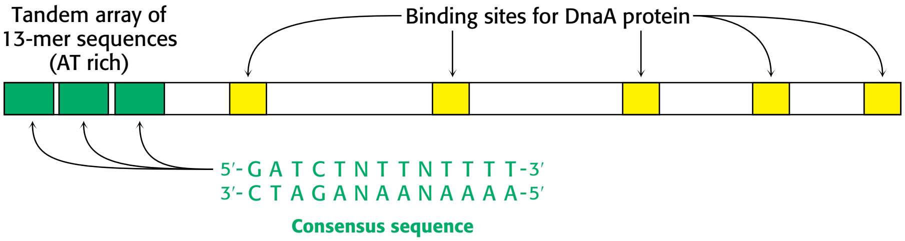

Tandem array of 13-mer sequences (AT rich)
Binding sites for DnaA protein
5′-G A T C T N T T N T T T T-3′
3′-C T A G A N A A N A A A A-5′
Consensus sequence

FIGURE 28.25 Origin of replication in *E. coli.* The *oriC* locus has a length of 245 bp. It contains a tandem array of three nearly identical 13-nucleotide sequences (green) and five binding sites (yellow) for the DnaA protein.

The gaps between fragments of the nascent lagging strand are filled by DNA polymerase I. This essential enzyme also uses its 5′ → 3′ exonuclease activity to remove the RNA primer lying ahead of the polymerase site. The primer cannot be erased by DNA polymerase III, because the enzyme lacks 5′ → 3′ editing capability. Finally, DNA ligase connects the fragments.

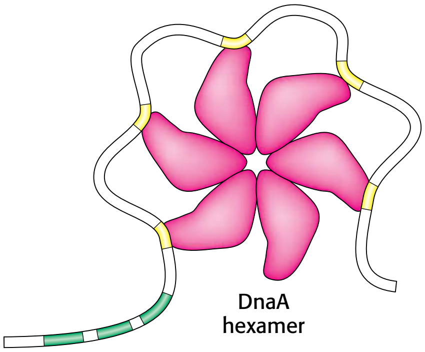


FIGURE 28.26 Assembly of DnaA. Monomers of DnaA bind to their binding sites (shown in yellow) in *oriC* and come together to form a complex structure, possibly the cyclic hexamer shown here. This structure marks the origin of replication and favors DNA strand separation in the AT-rich sites (green).

DNA replication in *Escherichia coli* begins at a unique site

In *E. coli,* DNA replication starts at a unique site within the entire 4.6 × 10^6 bp genome. This *origin of replication,* called the oriC *locus,* is a 245-bp region that has several unusual features (Figure 28.25). The *oriC* locus contains five copies of a sequence that is the preferred binding site for the origin-recognition protein DnaA. In addition, the locus contains a tandem array of 13-bp sequences that are rich in AT base pairs. Several steps are required to prepare for the start of replication:

1. *The binding of DnaA proteins to DNA is the first step in the preparation for replication.* DnaA is a member of the P-loop NTPase family related to the hexameric helicases. Each DnaA monomer comprises an ATPase domain linked to a DNA-binding domain at its C-terminus. DnaA molecules are able to bind to each other through their ATPase domains; a group of bound DnaA molecules will break apart on the binding and hydrolysis of ATP. The binding of DnaA molecules to one another signals the start of the preparatory phase, and their breaking apart signals the end of that phase. The DnaA proteins bind to the five high-affinity sites in *oriC* and then come together with DnaA molecules bound to lower-affinity sites to form an oligomer, possibly a cyclic hexamer. The DNA is wrapped around the outside of the DnaA hexamer (Figure 28.26).

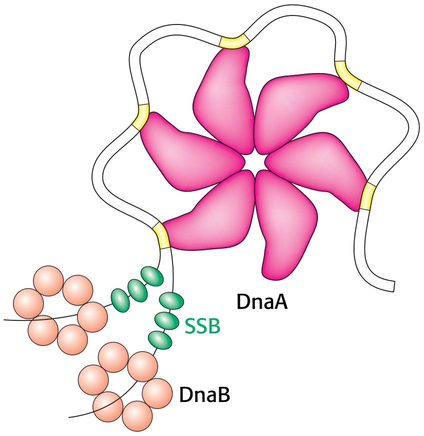


FIGURE 28.27 Prepriming complex. The AT-rich regions are unwound and trapped by the single-stranded-binding protein (SSB). The hexameric DNA helicase DnaB is loaded on each strand. At this stage, the complex is ready for the synthesis of the RNA primers and assembly of the DNA polymerase III holoenzyme.

2. *Single DNA strands are exposed in the prepriming complex.* With DNA wrapped around a DnaA hexamer, additional proteins are brought into play. The hexameric helicase DnaB is loaded around the DNA with the help of the helicase loader protein DnaC. Local regions of *oriC,* including the AT regions, are unwound and trapped by the single-stranded-DNA–binding protein. The result of this process is the generation of a structure called the *prepriming complex,* which makes single-stranded DNA accessible to other proteins (Figure 28.27). Significantly, the primase, DnaG, is now able to insert the RNA primer.

3. *The polymerase holoenzyme assembles.* The DNA polymerase III holoenzyme assembles on the prepriming complex, initiated by interactions between DnaB and the sliding-clamp subunit of DNA polymerase III. These interactions also trigger ATP hydrolysis within the DnaA subunits, signaling the initiation of DNA replication. The breakup of the DnaA

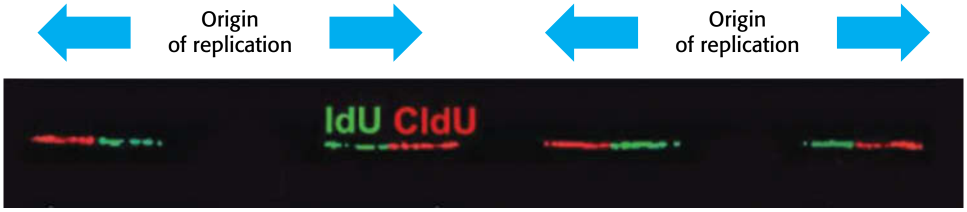


FIGURE 28.28 Eukaryotic origins of replication. The image shows a single molecule of DNA containing two origins of replication. The origins were identified by labeling new replicated DNA in human cells first with one thymine analog (iodo-deoxyuridine, I-dU) and then another (chloro-deoxyuridine, Cl-dU). DNA molecules from these cells were then extended on a microscope slide and labeled with antibodies to I-dU (green) and Cl-dU (red) to visualize the DNA. This method allows the detection of replication origins as well as determination of the rate of DNA synthesis. [Courtesy Aaron Bensimon. Data from C. Conti, et al., *Mol. Biol. Cell.* 18:3059-3067, 2007.]

assembly prevents additional rounds of replication from beginning at the replication origin.

DNA synthesis in eukaryotes is initiated at multiple sites

Replication in eukaryotes is mechanistically similar to replication in prokaryotes but is more challenging for a number of reasons. One of them is sheer size: *E. coli* must replicate 4.6 million base pairs, whereas a human diploid cell must replicate more than 6 billion base pairs. Second, the genetic information for *E. coli* is contained on 1 chromosome, whereas, in human beings, 23 pairs of chromosomes must be replicated. Finally, whereas the *E. coli* chromosome is circular, human chromosomes are linear. Unless countermeasures are taken, linear chromosomes are subject to shortening with each round of replication.

The first two challenges are met by the use of multiple origins of replication. In human beings, replication requires about 30,000 origins of replication, with each chromosome containing several hundred. Each origin of replication is the starting site for a replication unit, or *replicon.* DNA replication can be monitored by single-molecule methods, revealing bidirectional synthesis from different sites (Figure 28.28). In contrast with *E. coli,* the origins of replication in human beings do not contain regions of sharply defined sequence. Instead, more broadly defined AT-rich sequences are the sites around which the *origin of replication complexes* (ORCs) are assembled.

1. *The assembly of the ORC is the first step in the preparation for replication.* In human beings, the ORC is composed of six different proteins, each homologous to DnaA. These proteins come together to form a hexameric structure analogous to the assembly formed by DnaA.

2. *Licensing factors recruit a helicase that exposes single strands of DNA.* After the ORC has been assembled, additional proteins are recruited, including Cdc6, a homolog of the ORC subunits, and Cdt1. These proteins, in turn, recruit a hexameric helicase with six distinct subunits called Mcm2-7. These proteins, including the helicase, are sometimes called *licensing factors* because they permit the formation of the initiation complex. After the initiation complex has formed, Mcm2-7 separates the parental DNA strands, and the single strands are stabilized by the binding of *replication protein A,* a single-stranded-DNA–binding protein.

3. *Two distinct polymerases are needed to copy a eukaryotic replicon.* An initiator polymerase called *polymerase* α begins replication but is soon replaced by a more processive enzyme. This process is called *polymerase switching* because

TABLE 28.1 Some types of DNA polymerases

Name	Function
Prokaryotic Polymerases	
DNA polymerase I	Erases primer and fills in gaps on lagging strand
DNA polymerase II (error-prone polymerase)	DNA repair
DNA polymerase III	Primary enzyme of DNA synthesis
Eukaryotic Polymerases	
DNA polymerase α	Initiator polymerase
Primase subunit	Synthesizes the RNA primer
DNA polymerase unit	Adds stretch of about 20 nucleotides to the primer
DNA polymerase β (error-prone polymerase)	DNA repair
DNA polymerase δ	Primary enzyme of DNA synthesis

one polymerase has replaced another. This second enzyme, called *DNA polymerase* δ, is the principal replicative polymerase in eukaryotes (Table 28.1).

Replication begins with the binding of DNA polymerase α. This enzyme includes a primase subunit, used to synthesize the RNA primer, as well as an active DNA polymerase. After this polymerase has added a stretch of about 20 deoxynucleotides to the primer, another replication protein, called *replication factor C* (RFC), displaces DNA polymerase α. Replication factor C attracts a sliding clamp called *proliferating cell nuclear antigen* (PCNA), which is homologous to the β_2 subunit of *E. coli* polymerase III. The binding of PCNA to DNA polymerase δ renders the enzyme highly processive and suitable for long stretches of replication. Replication continues in both directions from the origin of replication until adjacent replicons meet and fuse. RNA primers are removed and the DNA fragments are ligated by DNA ligase.

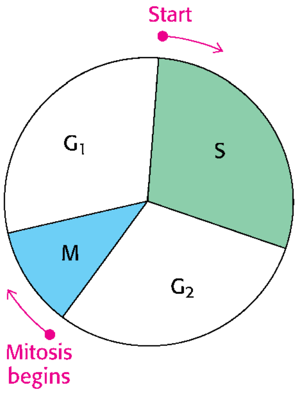


FIGURE 28.29 Eukaryotic cell cycle. DNA replication and cell division must take place in a highly coordinated fashion in eukaryotes. Mitosis (M) takes place only after DNA synthesis (S). Two gaps (G_1 and G_2) in time separate the two processes.

The use of multiple origins of replication requires mechanisms for ensuring that each sequence is replicated once and only once. The events of eukaryotic DNA replication are linked to the eukaryotic *cell cycle* (Figure 28.29). The processes of DNA synthesis and cell division are coordinated in the cell cycle so that the replication of all DNA sequences is complete before the cell progresses into the next phase of the cycle. This coordination requires several *checkpoints* that control the progression along the cycle. A family of small proteins termed *cyclins* are synthesized and degraded by proteasomal digestion in the course of the cell cycle. Cyclins act by binding to specific *cyclin-dependent protein kinases* and activating them. One such kinase, cyclin-dependent kinase 2 (cdk2) binds to assemblies at origins of replication and regulates replication through a number of interlocking mechanisms.

Telomeres are unique structures at the ends of linear chromosomes

Whereas the genomes of essentially all prokaryotes are circular, the chromosomes of human beings and other eukaryotes are linear. The free ends of linear DNA molecules introduce several complications that must be resolved by special enzymes. In particular, complete replication of DNA ends is difficult because polymerases act only in the $5' \rightarrow 3'$ direction. The lagging strand would have an incomplete 5′end after the removal of the RNA primer. Each round of replication would further shorten the chromosome.

The first clue to how this problem is resolved came from sequence analyses of the ends of chromosomes, which are called *telomeres* (from the Greek *telos,* "an end"). Telomeric DNA contains hundreds of tandem repeats of a six-nucleotide sequence. One of the strands is G rich at the 3′ end, and it is

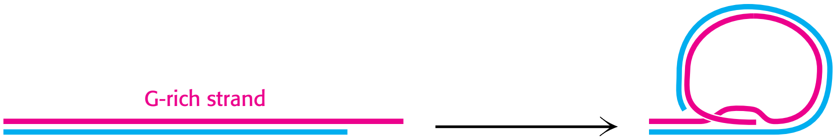


FIGURE 28.30 Proposed model for telomeres. A single-stranded segment of the G-rich strand extends from the end of the telomere. In one model for telomeres, this single-stranded region invades the duplex to form a large duplex loop.

slightly longer than the other strand. In human beings, the repeating G-rich sequence is AGGGTT.

The structure adopted by telomeres has been extensively investigated. Evidence suggests that they may form large duplex loops (Figure 28.30). The single-stranded region at the very end of the structure has been proposed to loop back to form a DNA duplex with another part of the repeated sequence, displacing a part of the original telomeric duplex. This loop-like structure is formed and stabilized by specific telomere-binding proteins. Such structures would nicely mask and protect the end of the chromosome.

Telomeres are replicated by telomerase, a specialized polymerase that carries its own RNA template

How are the repeated sequences generated? An enzyme, termed *telomerase,* that executes this function has been purified and characterized. When a primer ending in GGTT is added to human telomerase in the presence of deoxynucleoside triphosphates, the sequences GGTTAGGGTT and GGTTAGGGTTAGGGTT, as well as longer products, are generated. Elizabeth Blackburn and Carol Greider discovered that the enzyme adding the repeats contains an RNA molecule that serves as the template for the elongation of the G-rich strand (Figure 28.31). Thus, telomerase carries the information necessary to generate the telomere sequences. The exact number of repeated sequences is not crucial.

Subsequently, a protein component of telomerases also was identified. This component is related to reverse transcriptases, enzymes first discovered in retroviruses that copy RNA into DNA (Section 5.2). Thus, *telomerase is a specialized reverse transcriptase that carries its own template.* Telomerase is generally expressed at high levels only in rapidly growing cells. Thus, telomeres and telomerase can play important roles in cancer-cell biology and in cell aging.

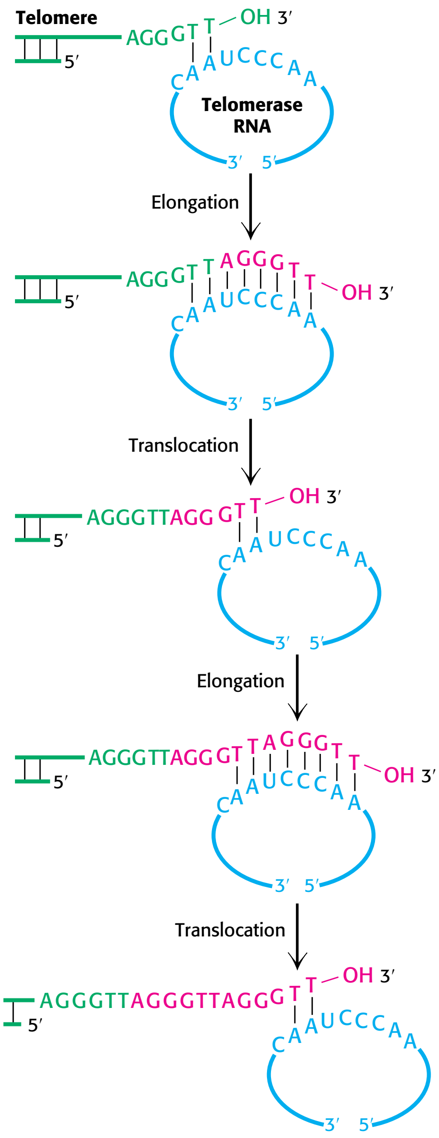


FIGURE 28.31 Telomere formation. Mechanism of synthesis of the G-rich strand of telomeric DNA. The RNA template of telomerase is shown in blue and the nucleotides added to the G-rich strand of the primer are shown in red. [Information from E. H. Blackburn, *Nature* 350:569–573, 1991.]

Because cancer cells express high levels of telomerase, whereas most normal cells do not, telomerase is a potential target for anticancer therapy. A variety of approaches for blocking telomerase expression or blocking its activity are under investigation for cancer treatment and prevention.

28.4 Many Types of DNA Damage Can Be Repaired

We have examined how even very large and complex genomes can, in principle, be replicated with considerable fidelity. However, DNA does become damaged, both in the course of replication and through other processes. Damage to DNA can be as simple as the misincorporation of a single base or it can take more-complex forms such as the chemical modification of bases, chemical cross-links between the two strands of the double helix, or breaks in one or both of the phosphodiester backbones. The results may be cell death or cell transformation, changes in the DNA sequence that can be inherited by future generations, or blockage of the DNA replication process itself. A variety of DNA-repair systems have evolved that can recognize these defects and, in many cases, restore the DNA molecule to its undamaged form. We begin with some of the sources of DNA damage.

Errors can arise in DNA replication

Errors introduced in the replication process are the simplest source of damage in the double helix. With the addition of each base, there is the possibility that an incorrect base might be incorporated, forming a non-Watson–Crick base pair. These non-Watson–Crick base pairs can locally distort the DNA double helix. Furthermore, such mismatches can be *mutagenic;* that is, they can result in permanent changes in the DNA sequence. When a double helix containing a non-Watson–Crick base pair is replicated, the two daughter double helices will have different sequences because the mismatched base is very likely to pair with its Watson–Crick partner. Errors other than mismatches include insertions, deletions, and breaks in one or both strands. Furthermore, replicative polymerases can stall or even fall off a damaged template entirely. As a consequence, replication of the genome may halt before it is complete.

A variety of mechanisms have evolved to deal with such interruptions, including specialized DNA polymerases that can replicate DNA across many lesions. A drawback is that such polymerases are substantially more error prone than are normal replicative polymerases. Nonetheless, these *translesion* or *error-prone polymerases* allow the completion of a draft sequence of the genome that can be at least partly repaired by DNA-repair processes. DNA recombination (Section 28.5) provides an additional mechanism for salvaging interruptions in DNA replication.

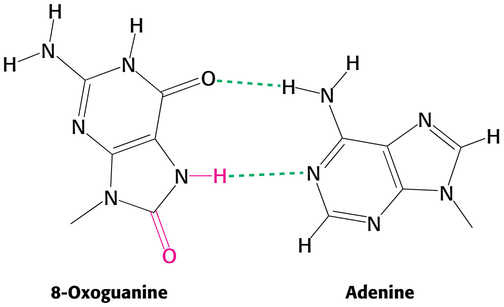


FIGURE 28.32 Oxoguanine–adenine base pair. When guanine is oxidized to 8-oxoguanine, the damaged base can form a base pair with adenine through an edge of the base that does not normally participate in base-pair formation.

Bases can be damaged by oxidizing agents, alkylating agents, and light

A variety of chemical agents can alter specific bases within DNA after replication is complete. Such *mutagens* include reactive oxygen species such as hydroxyl radical. For example, hydroxyl radical reacts with guanine to form 8-oxoguanine. 8-Oxoguanine is mutagenic because it often pairs with adenine rather than cytosine in DNA replication. Its choice of pairing partner differs from that of guanine because it uses a different edge of the base to form base pairs (Figure 28.32). Deamination is another potentially deleterious process. For example, adenine can be deaminated to form hypoxanthine (Figure 28.33). This process is mutagenic because hypoxanthine pairs with cytosine rather than thymine. Guanine and cytosine also can be deaminated to yield bases that pair differently from the parent base.

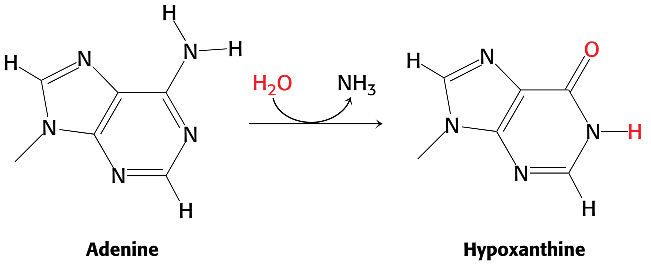
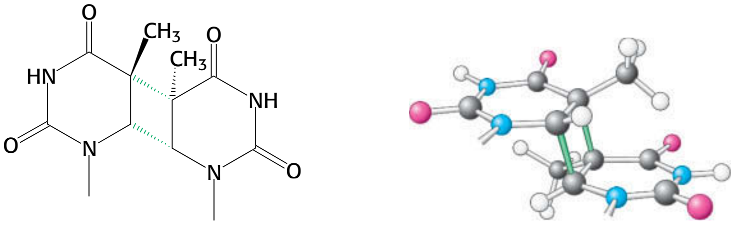


FIGURE 28.33 Adenine deamination. The base adenine can be deaminated to form hypoxanthine. Hypoxanthine forms base pairs with cytosine in a manner similar to that of guanine, and so the deamination reaction can result in mutation.

In addition to oxidation and deamination, nucleotide bases are subject to alkylation. Electrophilic centers can be attacked by nucleophiles such as N-7 of guanine and adenine to form alkylated adducts. Some compounds are converted into highly active electrophiles through the action of enzymes that normally play a role in detoxification. A striking example is aflatoxin B_1, a compound produced by molds that grow on peanuts and other foods. A cytochrome P450 enzyme (Section 26.4) converts this compound into a highly reactive epoxide (Figure 28.34). This agent reacts with the N-7 atom of guanosine to form a mutagenic adduct that frequently leads to a G–C-to-T–A transversion.

The ultraviolet component of sunlight is a ubiquitous DNA-damaging agent. Its major effect is to covalently link adjacent pyrimidine residues along a DNA strand (Figure 28.35). Such a pyrimidine dimer cannot fit into a double helix, and so replication and gene expression are blocked until the lesion is removed.



Thymine dimer

FIGURE 28.35 **Cross-linked dimer of two thymine bases.** Ultraviolet light induces cross-links between adjacent pyrimidines along one strand of DNA.

A thymine dimer is an example of an *intra*strand cross-link because both participating bases are in the same strand of the double helix. Cross-links between bases on opposite strands also can be introduced by various agents. Psoralens are compounds, produced by a number of plants including the common fig, that form such *inter*strand cross-links (Figure 28.36). Interstrand cross-links disrupt replication because they prevent strand separation.

High-energy electromagnetic radiation such as x-rays can damage DNA by producing high concentrations of reactive species in solution. X-ray exposure can induce several types of DNA damage including single- and double-stranded breaks in DNA. This ability to induce such DNA damage led Hermann Muller to discover the mutagenic effects of x-rays in *Drosophila* in 1927. This discovery contributed to the development of *Drosophila* as one of the premier organisms for genetic studies.

DNA damage can be detected and repaired by a variety of systems

To protect the genetic message, a wide range of DNA-repair systems are present in most organisms. Many systems repair DNA by using sequence information from the uncompromised strand. Such single-strand replication systems follow a similar mechanistic outline:

1. Recognize the offending base(s).
2. Remove the offending base(s).
3. Repair the resulting gap with a DNA polymerase and DNA ligase.

We will briefly consider examples of several repair pathways. Although many of these examples are taken from *E. coli*, corresponding repair systems are present in most other organisms, including humans.

The replicative DNA polymerases themselves are able to correct many DNA mismatches produced in the course of replication. For example, the ε subunit of *E. coli* DNA polymerase III functions as a 3′-to-5′ exonuclease. This domain removes mismatched nucleotides from the 3′ end of DNA by hydrolysis. How does the enzyme sense whether a newly added base is correct? As a new strand of DNA is synthesized, it is *proofread*. If an incorrect base is inserted, then DNA synthesis slows down owing to the difficulty of threading

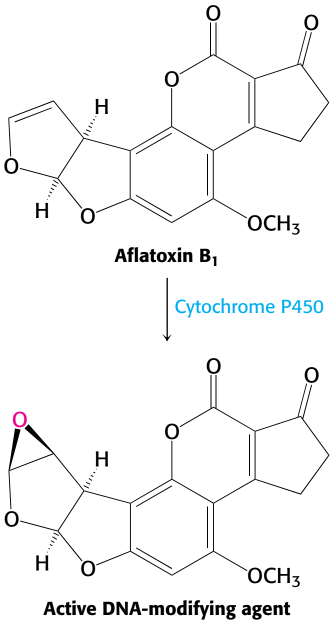
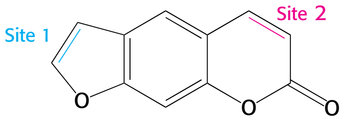


FIGURE 28.34 **Aflatoxin activation.** The compound, produced by molds that grow on peanuts, is activated by cytochrome P450 to form a highly reactive species that modifies bases such as guanine in DNA, leading to mutations.



Psoralen

FIGURE 28.36 **A cross-linking agent.** The compound psoralen and its derivatives can form interstrand cross-links through two reactive sites that can form adducts with nucleotide bases.

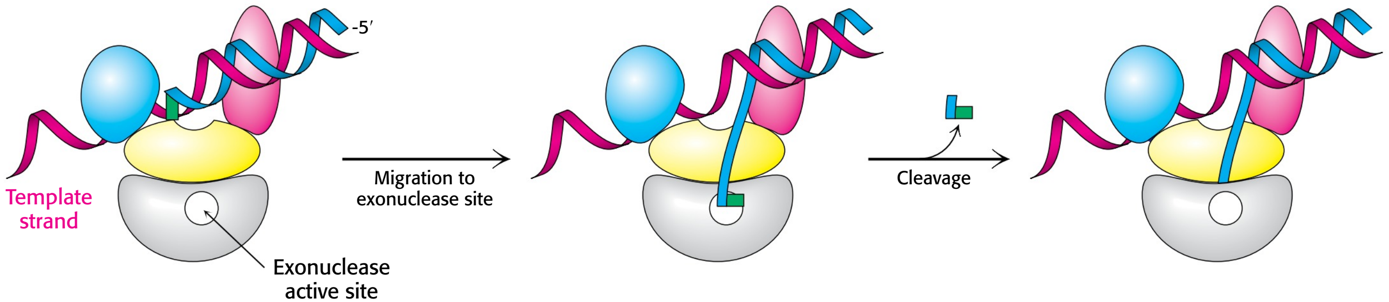


FIGURE 28.37 Proofreading. The growing polynucleotide chain occasionally leaves the polymerase site and migrates to the active site of the exonuclease. There, one or more nucleotides are excised from the newly synthesized chain, removing potentially incorrect bases.

a non-Watson–Crick base pair into the polymerase. In addition, the mismatched base is weakly bound and therefore able to fluctuate in position. The delay from the slowdown allows time for these fluctuations to take the newly synthesized strand out of the polymerase active site and into the exonuclease active site (Figure 28.37). There, the DNA is degraded, one nucleotide at a time, until it moves back into the polymerase active site and synthesis continues.

A second mechanism is present in essentially all cells to correct errors made in replication that are not corrected by proofreading (Figure 28.38). *Mismatch-repair* systems consist of at least two proteins, one for detecting the mismatch and the other for recruiting an endonuclease that cleaves the newly synthesized DNA strand close to the lesion to facilitate repair. In *E. coli,* these proteins are called MutS and MutL and the endonuclease is called MutH.

Another mechanism of DNA repair is *direct repair,* one example of which is the photochemical cleavage of pyrimidine dimers. Nearly all cells contain a *photoreactivating enzyme* called *DNA photolyase.* The *E. coli* enzyme, a 35-kd protein that contains bound N^{5},N^{10}-methenyltetrahydrofolate and flavin adenine dinucleotide (FAD) cofactors, binds to the distorted region of DNA. The enzyme uses light energy—specifically, the absorption of a photon by the N^{5},N^{10}-methenyltetrahydrofolate coenzyme—to form an excited state that cleaves the dimer into its component bases.

The excision of modified bases such as 3-methyladenine by the *E. coli* enzyme *AlkA* is an example of *base-excision repair.* The binding of this enzyme to damaged DNA flips the affected base out of the DNA double helix and into the active site of the enzyme (Figure 28.39). The enzyme then

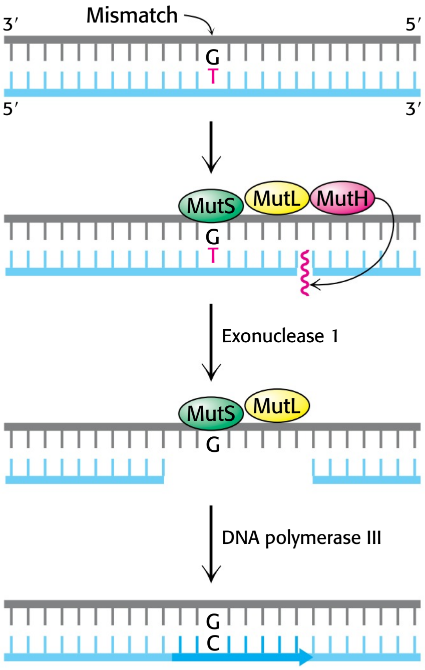


FIGURE 28.38 Mismatch repair. DNA mismatch repair in *E. coli* is initiated by the interplay of MutS, MutL, and MutH proteins. A G—T mismatch is recognized by MutS. MutH cleaves the backbone in the vicinity of the mismatch. A segment of the DNA strand containing the erroneous T is removed by exonuclease I and synthesized anew by DNA polymerase III. [Information from R. F. Service, *Science* 263:1559–1560, 1994.]

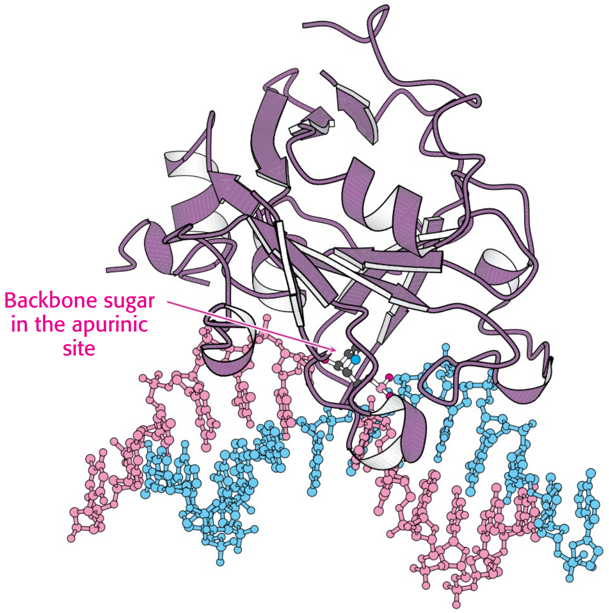


FIGURE 28.39 Structure of DNA-repair enzyme. A complex between the DNA-repair enzyme AlkA and an analog of a DNA molecule missing a purine base (an apurinic site) is shown. *Notice* that the backbone sugar in the apurinic site is flipped out of the double helix into the active site of the enzyme. [Drawn from 1BNK.pdb.]

acts as *a glycosylase,* cleaving the glycosidic bond to release the damaged base. At this stage, the DNA backbone is intact, but a base is missing. This hole is called an *AP site* because it is apurinic (devoid of A or G) or apyrimidinic (devoid of C or T). An *AP endonuclease* recognizes this defect and nicks the backbone adjacent to the missing base. *Deoxyribose phosphodiesterase* excises the residual deoxyribose phosphate unit, and DNA polymerase I inserts an undamaged nucleotide, as dictated by the base on the undamaged complementary strand. Finally, the repaired strand is sealed by DNA ligase.

One of the best-understood examples of *nucleotide-excision repair* is utilized for the excision of a pyrimidine dimer. Three enzymatic activities are essential for this repair process in *E. coli* (Figure 28.40). First, an enzyme complex consisting of the proteins encoded by the *uvrABC* genes detects the distortion produced by the DNA damage. The UvrABC enzyme then cuts the damaged DNA strand at two sites, 8 nucleotides away from the damaged site on the 5′ side and 4 nucleotides away on the 3′ side. The 12-residue oligonucleotide excised by this highly specific *excinuclease* (from the Latin *exci,* "to cut out") then diffuses away. DNA polymerase I enters the gap to carry out repair synthesis. The 3′ end of the nicked strand is the primer, and the intact complementary strand is the template. Finally, the 3′ end of the newly synthesized stretch of DNA and the original part of the DNA chain are joined by DNA ligase.

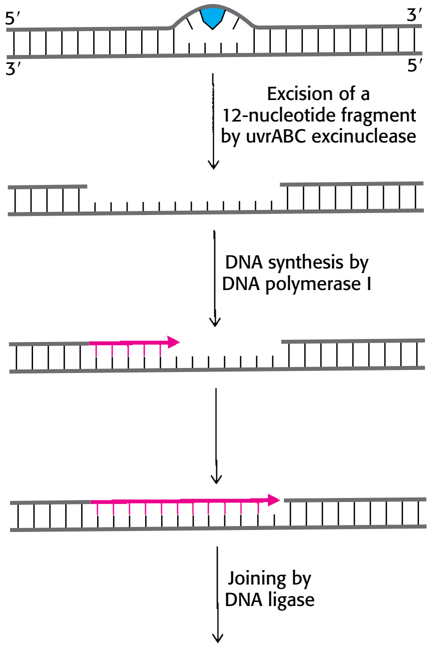


FIGURE 28.40 Nucleotide-excision repair. Repair of a region of DNA containing a thymine dimer by the sequential action of a specific excinuclease, a DNA polymerase, and a DNA ligase. The thymine dimer is shown in blue and the new region of DNA is in red. [Information from P. C. Hanawalt, *Endevour* 31:83, 1982.]

DNA ligase is able to seal simple breaks in one strand of the DNA backbone. However, alternative mechanisms are required to repair breaks on both strands that are close enough together to separate the DNA into two double helices. Several distinct mechanisms are able to repair such damage. One mechanism, *nonhomologous end joining* (NHEJ), does not depend on other DNA molecules in the cell. In NHEJ, the free double-stranded ends are bound by a heterodimer of two proteins, Ku70 and Ku80. These proteins stabilize the ends and mark them for subsequent manipulations. Through mechanisms that are not yet well understood, the Ku70/80 heterodimers act as handles used by other proteins to draw the two double-stranded ends close together so that enzymes can seal the break.

Alternative mechanisms of double-stranded-break repair can operate if an intact stretch of double-stranded DNA with an identical or very similar sequence is present in the cell. These repair processes use homologous recombination, presented in Section 28.5.

As noted in Section 5.4, researchers can now take advantage of double-stranded-break repair mechanisms for executing targeted changes in eukaryotic genomes. Specific engineered nucleases are used to introduce double-strand breaks in particular locations within the genome. The DNA repair machinery in the targeted cells will then repair the break, through either NHEJ to disrupt the gene or homologous recombination-based repair with added double-stranded DNA fragments to introduce more elaborate modifications. For example, mutations associated with human diseases can be specifically introduced into the genomes of model organisms such as mice or zebrafish to examine the consequences.

The presence of thymine instead of uracil in DNA permits the repair of deaminated cytosine

The presence in DNA of thymine rather than uracil, as in RNA, was an enigma for many years. Both bases pair with adenine. The only difference between them is a methyl group in thymine in place of the C-5 hydrogen atom in uracil. Why is a methylated base employed in DNA and not in RNA? The existence of an active repair system to correct the deamination of cytosine provides a convincing solution to this puzzle.

Cytosine in DNA spontaneously deaminates at a perceptible rate to form uracil. The deamination of cytosine is potentially mutagenic because uracil pairs with adenine, and so one of the daughter strands will contain a U–A base pair rather than the original C–G base pair. This mutation is prevented by a repair system that recognizes uracil to be foreign to DNA (Figure 28.41). The repair enzyme, *uracil DNA glycosylase,* is homologous to AlkA. The enzyme hydrolyzes the glycosidic bond between the uracil and deoxyribose moieties but does not attack thymine-containing nucleotides. The AP site generated is repaired to reinsert cytosine. Thus, *the methyl group on thymine is a tag that distinguishes thymine from deaminated cytosine.* If thymine were not used in DNA, uracil correctly in place would be indistinguishable from uracil formed by deamination. The defect would persist unnoticed, and so a C–G base pair would necessarily be mutated to U–A in one of the daughter DNA molecules. This mutation is prevented by a repair system that searches for uracil and leaves thymine alone. *Thymine is used instead of uracil in DNA to enhance the fidelity of the genetic message.*

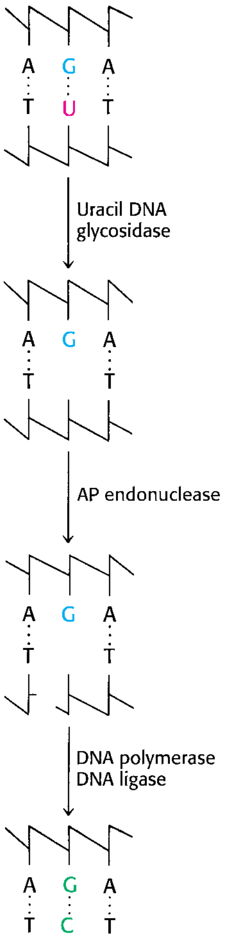


FIGURE 28.41 Uracil repair. Uridine bases in DNA, formed by the deamination of cytidine, are excised and replaced by cytidine.

Some genetic diseases are caused by the expansion of repeats of three nucleotides

Some genetic diseases are caused by the presence of DNA sequences that are inherently prone to errors in the course of repair and replication. A particularly important class of such diseases is characterized by the presence of long tandem arrays of repeats of three nucleotides. An example is *Huntington disease,* an autosomal dominant neurological disorder with a variable age of onset. The mutated gene in this disease expresses a protein in the brain called huntingtin, which contains a stretch of consecutive glutamine residues. These glutamine residues are encoded by a tandem array of CAG sequences within the gene. In unaffected persons, this array is between 6 and 31 repeats, whereas, in those with the disease, the array is between 36 and 82 repeats or longer. Moreover, the array tends to become longer from one generation to the next. The consequence is a phenomenon called *anticipation:* the children of an affected parent tend to show symptoms of the disease at an earlier age than did the parent.

The tendency of these *trinucleotide repeats* to expand is explained by the formation of alternative structures in the course of DNA repair. On cleavage of the DNA backbone, part of the array can loop out without disrupting base-pairing outside this region. Then, in replication, DNA polymerase extends this strand through the remainder of the array, leading to an increase in the number of copies of the trinucleotide sequence.

A number of other neurological diseases are characterized by expanding arrays of trinucleotide repeats. How do these long stretches of repeated amino acids cause disease? For huntingtin, it appears that the polyglutamine stretches become increasingly prone to aggregate as their length increases; the additional consequences of such aggregation are still under investigation.

Many cancers are caused by the defective repair of DNA

As described in Chapter 14, cancers are caused by mutations in genes associated with growth control. Defects in DNA-repair systems increase the overall frequency of mutations and, hence, the likelihood of cancer-causing mutations. Indeed, the synergy between studies of mutations that predispose people to cancer and studies of DNA repair

TABLE 28.2

Disease	Selected Genes	Repair Pathway
Xeroderma pigmentosum (skin)	XPA, XPB, XPC	Nucleotide excision repair
Lynch syndrome (colon cancer)	MSH2, MLH1	Mismatch repair
Breast and ovarian cancer	BRCA1 and BRCA2	Double-strand break repair
Renal and lung cancer	OGG1	Base excision repair

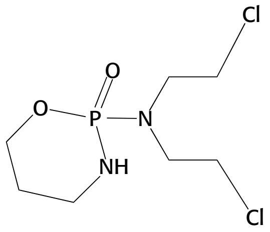
in model organisms has been tremendous in revealing the biochemistry of DNA-repair pathways. Genes for DNA-repair proteins are often *tumor suppressor genes;* that is, they suppress tumor development when at least one copy of the gene is free of a deleterious mutation. When both copies of a gene are mutated, however, tumors develop at rates greater than those for the population at large. People who inherit defects in a single tumor-suppressor allele do not necessarily develop cancer but are susceptible to developing the disease because only the one remaining normal copy of the gene must develop a new defect to further the development of cancer. Table 28.2 lists selected genes that are associated with diseases including cancer and that encode proteins involved with DNA repair processes.

Consider, for example, *xeroderma pigmentosum,* a rare human skin disease. The skin of an affected person is extremely sensitive to sunlight or ultraviolet light. In infancy, severe changes in the skin become evident and worsen with time. The skin becomes dry, and there is a marked atrophy of the dermis. Keratoses appear, the eyelids become scarred, and the cornea ulcerates. Skin cancer usually develops at several sites. Many patients die before age 30 from metastases of these malignant skin tumors. Studies of xeroderma pigmentosum patients have revealed that mutations occur in genes for a number of different proteins. These proteins are components of the human nucleotide-excision-repair pathway, including homologs of the UvrABC subunits.

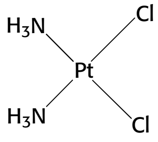
Defects in other repair systems can increase the frequency of other tumors. For example, *hereditary nonpolyposis colorectal cancer* (HNPCC, or *Lynch syndrome*) results from defective DNA mismatch repair. HNPCC is not rare—as many as 1 in 200 people will develop this form of cancer. Mutations in two genes, called *hMSH2* and *hMLH1,* account for most cases of this hereditary predisposition to cancer. The striking finding is that these genes encode the human counterparts of MutS and MutL of *E. coli*. Mutations in *hMSH2* and *hMLH1* seem likely to allow mutations to accumulate throughout the genome. In time, genes important in controlling cell proliferation become altered, resulting in the onset of cancer.

Not all tumor-suppressor genes are specific to particular types of cancer. *The gene for a protein called p53 is mutated in more than half of all tumors.* The p53 protein helps control the fate of damaged cells. First, it plays a central role in sensing DNA damage, especially double-stranded breaks. Then, after sensing damage, the protein either promotes a DNA-repair pathway or activates the apoptosis pathway, leading to cell death. Most mutations in the p53 gene are sporadic; that is, they occur in somatic cells rather than being inherited. People who inherit a deleterious mutation in one copy of the p53 gene suffer from *Li-Fraumeni syndrome* and have a high probability of developing several types of cancer.

Cancer cells often have two characteristics that make them especially vulnerable to agents that damage DNA molecules. First, they divide frequently, and so their DNA replication pathways are more active than they



Cyclophosphamide



Cisplatin

are in most cells. Second, as already noted, cancer cells often have defects in DNA-repair pathways. Several agents widely used in cancer chemotherapy, including cyclophosphamide and cisplatin, act by damaging DNA. Cancer cells are less able to avoid the effect of the induced damage than are normal cells, providing a therapeutic window for specifically killing cancer cells.

Many potential carcinogens can be detected by their mutagenic action on bacteria

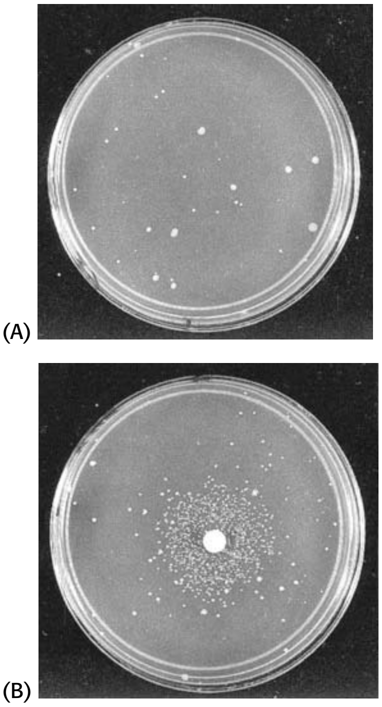


FIGURE 28.42 Ames test. (A) A petri plate containing about 10^9 *Salmonella* bacteria that cannot synthesize histidine and (B) a petri plate containing a filter-paper disc with a mutagen, which produces a large number of revertants that can synthesize histidine. After 2 days, the revertants appear as rings of colonies around the disc. The small number of visible colonies in plate A are spontaneous revertants. [Reprinted from *Mutation Research/Environmental Mutagenesis and Related Subjects*, 31:6, Ames, B.N., et al. Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test, p. 347–363, Copyright 1975, with permission from Elsevier.]

Many human cancers are caused by exposure to chemicals that cause mutations. It is important to identify such compounds and to ascertain their potency so that human exposure to them can be minimized. Bruce Ames devised a simple and sensitive test for detecting chemical mutagens. In the *Ames test,* a thin layer of agar containing about 10^9 bacteria of a specially constructed tester strain of *Salmonella* is placed on a petri plate. These bacteria are unable to grow in the absence of histidine, because a mutation is present in one of the genes for the biosynthesis of this amino acid. The addition of a chemical mutagen to the center of the plate results in many new mutations. A small proportion of them reverse the original mutation, and histidine can be synthesized. These *revertants* multiply in the absence of an external source of histidine and appear as discrete colonies after the plate has been incubated at 37°C for 2 days (Figure 28.42). For example, 0.5 μg of 2-aminoanthracene gives 11,000 revertant colonies, compared with only 30 spontaneous revertants in its absence. A series of concentrations of a chemical can be readily tested to generate a dose–response curve.

Some of the tester strains are responsive to *base-pair substitutions,* whereas others detect *deletions or additions of base pairs (frameshifts).* The sensitivity of these specially designed strains has been enhanced by the genetic deletion of their excision-repair systems. Potential mutagens enter the tester strains easily because the lipopolysaccharide barrier that normally coats the surface of *Salmonella* is incomplete in these strains. A key feature of this detection system is the inclusion of a *mammalian liver homogenate*. Recall that some potential carcinogens such as aflatoxin are converted into their active forms by enzyme systems in the liver or other mammalian tissues. Bacteria lack these enzymes, and so the test plate requires a few milligrams of a liver homogenate to activate this group of mutagens.

The *Salmonella* test is extensively used to help evaluate the mutagenic and carcinogenic risks of a large number of chemicals. This rapid and inexpensive bacterial assay for mutagenicity complements epidemiological surveys and animal tests that are necessarily slower, more laborious, and far more expensive. The *Salmonella* test for mutagenicity is an outgrowth of studies of gene–protein relations in bacteria. It is a striking example of how fundamental research in molecular biology can lead directly to important advances in public health.

28.5 DNA Recombination Plays Important Roles in Replication, Repair, and Other Processes

Most processes associated with DNA replication function to copy the genetic message as faithfully as possible. However, several biochemical processes require the *recombination* of genetic material between two DNA molecules. In genetic recombination, two daughter molecules are formed by the exchange of genetic material between two parent molecules (Figure 28.43). Recombination is essential in the following processes.

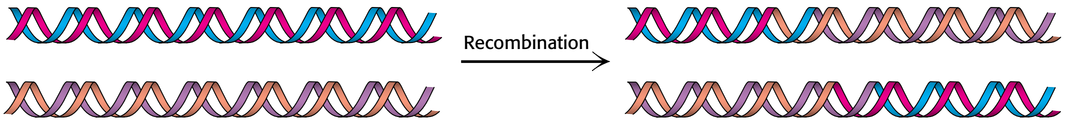


FIGURE 28.43 Recombination. Two DNA molecules can recombine with each other to form new DNA molecules that have segments from both parent molecules.

1. When replication stalls, recombination processes can reset the replication machinery so that replication can continue.

2. Some double-stranded breaks in DNA are repaired by recombination.

3. In meiosis, the limited exchange of genetic material between paired chromosomes provides a simple mechanism for generating genetic diversity in a population.

4. As we shall see in Chapter 34, recombination plays a crucial role in generating molecular diversity for antibodies and some other molecules in the immune system.

5. Some viruses employ recombination pathways to integrate their genetic material into the DNA of a host cell.

6. Recombination is used to manipulate genes in, for example, the generation of "gene knockout" mice and other targeted genome modifications (Section 5.4).

Recombination is most efficient between DNA sequences that are similar in sequence. In homologous recombination, parent DNA duplexes align at regions of sequence similarity, and new DNA molecules are formed by the breakage and joining of homologous segments.

RecA can initiate recombination by promoting strand invasion

In many recombination pathways, a DNA molecule with a free end recombines with a DNA molecule having no free ends available for interaction. DNA molecules with free ends are the common result of double-stranded DNA breaks, but they may also be generated in DNA replication if the replication complex stalls. This type of recombination has been studied extensively in *E. coli,* but it also takes place in other organisms through the action of proteins homologous to those of *E. coli*. Often dozens of proteins participate in the complete recombination process. However, the key protein is *RecA*, the homolog of which is called RAD51 in human cells. To accomplish the exchange, the single-stranded DNA displaces one of the strands of the double helix (Figure 28.44). The resulting three-stranded structure is called a *displacement loop* or *D-loop*. This process is often referred to as *strand invasion*. Because a free 3′ end is now base-paired to a contiguous strand of DNA, the 3′ end can act as a primer to initiate new DNA synthesis. Strand invasion can initiate many processes, including the repair of double-stranded breaks and the reinitiation of replication after the replication apparatus has come off its template. In the repair of a break, the recombination partner is an intact DNA molecule with an overlapping sequence.

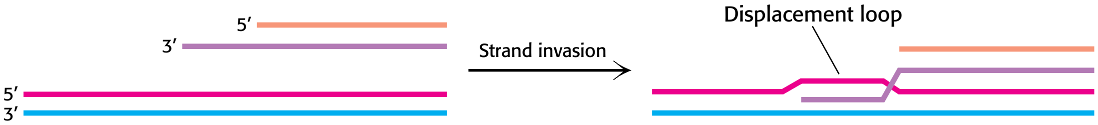


FIGURE 28.44 Strand invasion. This process, promoted by proteins such as RecA, can initiate recombination.

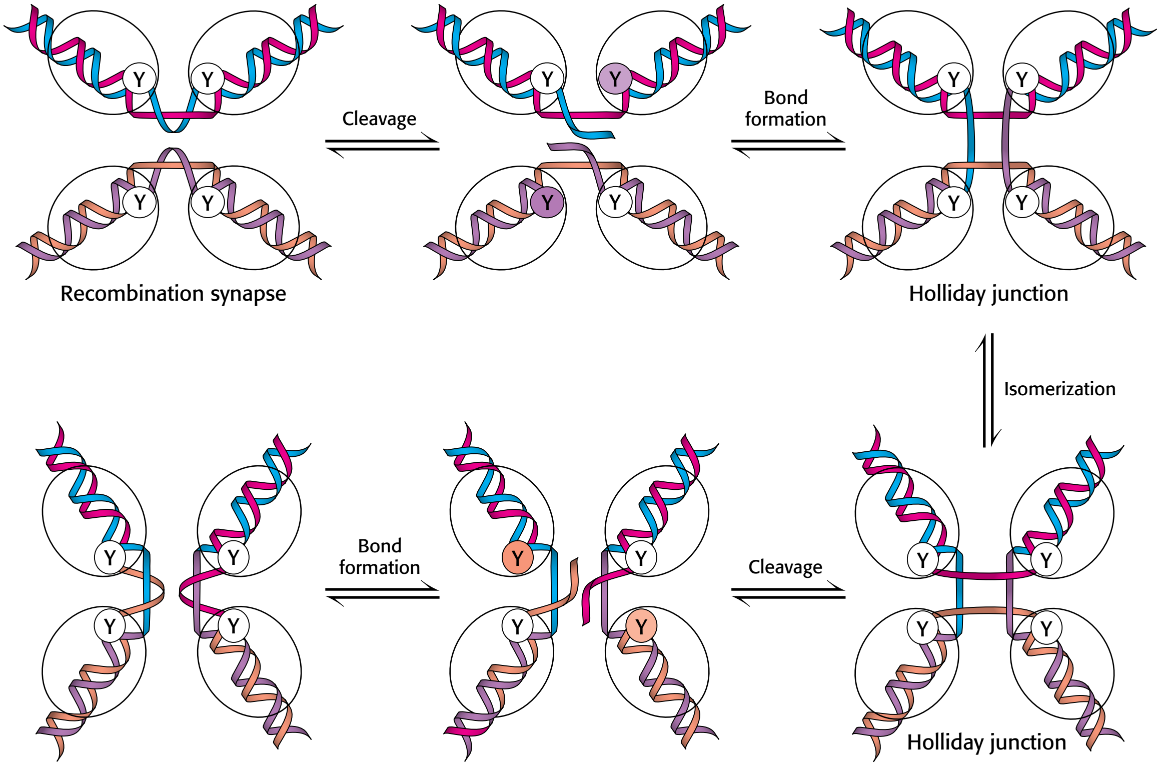


FIGURE 28.45 Recombination mechanism. Recombination begins as two DNA molecules come together to form a recombination synapse. One strand from each duplex is cleaved by transesterification with the 3′ end of each of the cleaved strands linked to a tyrosine residue on the recombinase enzyme. New phosphodiester bonds are formed when a 5′ end of the other cleaved strand in the complex attacks these tyrosine–DNA adducts. After isomerization, these steps are repeated to form the recombined products.

Some recombination reactions proceed through Holliday-junction intermediates

In recombination pathways for meiosis and some other processes, intermediates form that are composed of four polynucleotide chains in a crosslike structure. Intermediates with these crosslike structures are often referred to as *Holliday junctions,* after Robin Holliday, who proposed their role in recombination in 1964. Such intermediates have been characterized by a wide range of techniques, including x-ray crystallography.

Specific enzymes, termed *recombinases,* bind to these structures and resolve them into separated DNA duplexes. The Cre recombinase from bacteriophage P1 has been extensively studied. The mechanism begins with the recombinase binding to the DNA substrates (Figure 28.45).

Four molecules of the enzyme and two DNA molecules come together to form a *recombination synapse.* The reaction begins with the cleavage of one strand from each duplex via transesterification reactions with the 3′-phosphoryl groups becoming linked to specific tyrosine residues in the recombinase while the 5′-hydroxyl group of each cleaved strand remains free. The free 5′ ends invade the other duplex and participate in transesterification reactions to form new phosphodiester bonds and free the tyrosine residues. These reactions result in the formation of a Holliday junction. This junction can then isomerize to form a structure in which the polynucleotide chains in the center of the structure are reoriented. From this junction, the processes of strand cleavage and phosphodiester-bond formation repeat. The result is a synapse containing the two recombined duplexes. Dissociation of this complex generates the final recombined products.

Cre catalyzes the formation of Holliday junctions as well as their resolution. In contrast, other proteins bind to Holliday junctions that have already been formed by other processes and resolve them into separate duplexes. In many cases, these proteins also promote the process of branch migration whereby a Holliday junction is moved along the two component double helices. Branch migration can affect which segments of DNA are exchanged in a recombination process.

Recombination can also occur during cell division between different, non-homologous chromosomes. These recombination events, termed translocations, can be harmless or can result in diseases, depending on the chromosomes involved and the genes that are modified or disrupted. The formation of the bcr-abl gene, which leads to chronic myelogenous leukemia, was discussed in section 14.5.

SUMMARY

28.1 DNA Replication Proceeds by the Polymerization of Deoxyribonucleoside Triphosphates Along a Template

DNA polymerases are template-directed enzymes that catalyze the formation of phosphodiester bonds by nucleophilic attack by the 3′-hydroxyl group on the innermost phosphorus atom of a deoxyribonucleoside 5′-triphosphate. The complementarity of shape between correctly matched nucleotide bases is crucial to ensuring the fidelity of base incorporation. DNA polymerases cannot start chains de novo; a primer with a free 3′-hydroxyl group is required. Thus, DNA synthesis is initiated by the synthesis of an RNA primer, the task of a specialized primase enzyme. After serving as a primer, the RNA is degraded and replaced by DNA. DNA polymerases always synthesize a DNA strand in the 5′-to-3′ direction. So that both strands of the double helix can be synthesized in the same direction simultaneously, one strand is synthesized continuously while the other is synthesized in fragments called Okazaki fragments. Gaps between the fragments are sealed by DNA ligases. ATP-driven helicases prepare the way for DNA replication by separating the strands of the double helix.

28.2 DNA Unwinding and Supercoiling Are Controlled by Topoisomerases

A key topological property of DNA is its linking number *(Lk),* which is defined as the number of times one strand of DNA winds around the other in the right-hand direction when the DNA axis is constrained to lie in a plane. Molecules differing in linking number are topoisomers of one another and can be interconverted only by cutting one or both DNA strands; these reactions are catalyzed by topoisomerases. Changes in linking number generally lead to changes in both the number of turns of double helix and the number of turns of superhelix. Topoisomerase II catalyzes the ATP-driven introduction of negative supercoils, which leads to the compaction of DNA and renders it more susceptible to unwinding. Supercoiled DNA can be relaxed by topoisomerase I or topoisomerase II. Topoisomerase I acts by transiently cleaving one strand of DNA in a double helix, whereas topoisomerase II transiently cleaves both strands simultaneously.

28.3 DNA Replication Is Highly Coordinated

Replicative DNA polymerases are processive; that is, they catalyze the addition of many nucleotides without dissociating from the template. A major contributor to processivity is the DNA sliding clamp, such as the dimeric β subunit of the *E. coli* replicative polymerase. The sliding clamp has a ring structure that encircles the DNA double helix and keeps the enzyme and DNA associated. The DNA polymerase holoenzyme is a large DNA-copying machine formed by two DNA polymerase enzymes, one to act on each template strand, associated with other subunits including a sliding clamp and a clamp loader.

The synthesis of the leading and lagging strands of a double-stranded DNA template is coordinated. As a replicative polymerase moves along a DNA template, the leading strand is copied smoothly while the lagging strand forms loops that change length in the course of the synthesis of each Okazaki fragment. The mode of action is referred to as the trombone model.

DNA replication is initiated at a single site within the *E. coli* genome. A set of specific proteins recognizes this origin of replication and assembles the enzymes needed for DNA synthesis, including a helicase that promotes strand separation. The initiation of replication in eukaryotes is more complex. DNA synthesis is initiated at thousands of sites throughout the genome. Complexes homologous to those in *E. coli,* but more complicated, are assembled at each eukaryotic origin of replication. A special polymerase called telomerase that relies on an RNA template synthesizes specialized structures called telomeres at the ends of linear chromosomes.

28.4 Many Types of DNA Damage Can Be Repaired

A wide variety of DNA damage can occur. For example, mismatched bases may be incorporated in the course of DNA replication or individual bases may be damaged by oxidation or alkylation after DNA replication. Other forms of damage are the formation of cross-links and the introduction of single- or double-stranded breaks in the DNA backbone. Several different repair systems detect and repair DNA damage. Repair begins with the process of proofreading in DNA replication: mismatched bases that were incorporated in the course of synthesis are excised by exonuclease activity present in replicative polymerases. Some DNA lesions such as thymine dimers can be directly reversed through the action of specific enzymes. Other DNA-repair pathways act through the excision of single damaged bases (base-excision repair) or short segments of nucleotides (nucleotide-excision repair). Double-stranded breaks in DNA can be repaired by homologous or nonhomologous end-joining processes. Defects in DNA-repair components are associated with susceptibility to many different sorts of cancer. Such defects are a common target of cancer treatments. Many potential carcinogens can be detected by their mutagenic action on bacteria (the Ames test).

28.5 DNA Recombination Plays Important Roles in Replication, Repair, and Other Processes

Recombination is the exchange of segments between two DNA molecules. Recombination is important in some types of DNA repair as well as other processes such as meiosis, the generation of antibody diversity, and the life cycles of some viruses. Some recombination pathways are initiated by strand invasion, in which a single strand at the end of a DNA double helix forms base pairs with one strand of DNA in another double helix and displaces the other strand. A common intermediate formed in other recombination pathways is the Holliday junction, which consists of four strands of DNA that come together to form a crosslike structure. Recombinases promote recombination reactions through the introduction of specific DNA breaks and the formation and resolution of Holliday-junction intermediates.

KEY TERMS

template (p. 827)
DNA polymerase (p. 829)
primer (p. 829)
exonuclease (p. 829)
primase (p. 831)
replication fork (p. 831)
Okazaki fragment (p. 831)
lagging strand (p. 832)
leading strand (p. 832)

DNA ligase (p. 832)
helicase (p. 832)
supercoiling (p. 834)
linking number (p. 835)
topoisomer (p. 835)
twist (p. 835)
writhe (p. 835)
topoisomerase (p. 836)
processivity (p. 839)
sliding clamp (p. 840)
trombone model (p. 841)
origin of replication (p. 842)
origin of replication complex (ORC) (p. 843)
cell cycle (p. 844)
telomere (p. 844)
telomerase (p. 845)
mutagen (p. 846)
mismatch repair (p. 848)
direct repair (p. 848)
base-excision repair (p. 848)
nucleotide-excision repair (p. 849)
nonhomologous end joining (NHEJ) (p. 849)
trinucleotide repeat (p. 850)
tumor-suppressor gene (p. 851)
Ames test (p. 852)
RecA (p. 853)
Holliday junction (p. 854)
recombinase (p. 854)
recombination synapse (p. 854)

PROBLEMS

1. *Activated intermediates.* DNA polymerase I, DNA ligase, and topoisomerase I catalyze the formation of phosphodiester bonds. What is the activated intermediate in the linkage reaction catalyzed by each of these enzymes? What is the leaving group?

2. *Life in a hot tub.* An archaeon *(Sulfolobus acidocaldarius)* found in acidic hot springs contains a topoisomerase that catalyzes the ATP-driven introduction of positive supercoils into DNA. How might this enzyme be advantageous to this unusual organism?

3. *Which way?* Provide a chemical explanation of why DNA synthesis proceeds in a 5′-to-3′ direction.

4. *Nucleotide requirement*. DNA replication does not take place in the absence of the ribonucleotides ATP, CTP, GTP, and UTP. Propose an explanation.

5. *Close contact*. Examination of the structure of DNA polymerases bound to nucleotide analogs reveals that conserved residues come within van der Waals contact of C-2′ of the bound nucleotide. What is the potential significance of this interaction?

6. *Molecular motors in replication.*

(a) How fast does template DNA spin (expressed in revolutions per second) at an *E. coli* replication fork?

(b) What is the velocity of movement (in micrometers per second) of DNA polymerase III holoenzyme relative to the template?

7. *Wound tighter than a drum.* Why would replication come to a halt in the absence of topoisomerase II?

8. *The missing link*. One form of a plasmid shows a twist of $Tw = 48$ and a writhe of $Wr = 3$.

(a) What is the linking number?

(b) What would the value of writhe be for a form with twist $Tw = 50$ if the linking number is the same as that for the preceding form?

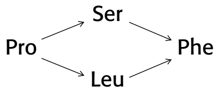
9. *Telomeres and cancer.* Telomerase is not active in most human cells. Some cancer biologists have suggested that activation of the telomerase gene would be a requirement for a cell to become cancerous. Explain why this might be the case.

10. *Backward?* Bacteriophage T7 helicase moves along DNA in the 5′-to-3′ direction. Other helicases have been reported to move in the 3′-to-5′ direction. Is there any fundamental reason why you would expect helicases to move in one direction or the other?

11. *Nick translation*. Suppose that you wish to make a sample of DNA duplex highly radioactive to use as a DNA probe. You have a DNA endonuclease that cleaves the DNA internally to generate 3′-OH and 5′-phosphoryl groups, intact DNA polymerase I, and radioactive dNTPs. Suggest a means for making the DNA radioactive.

12. *Revealing tracks.* Suppose that replication is initiated in a medium containing *moderately* radioactive tritiated thymine. After a few minutes of incubation, the bacteria are transferred to a medium containing *highly* radioactive tritiated thymidine. Sketch the autoradiographic pattern that would be seen for (a) undirectional replication and (b) bidirectional replication, each from a single origin.

13. *Mutagenic trail.* Suppose that the single-stranded RNA from tobacco mosaic virus was treated with a chemical mutagen, that mutants were obtained having serine or leucine instead of proline at a specific position, and that further treatment of these mutants with the same mutagen yielded phenylalanine at this position.



What are the plausible codon assignments for these four amino acids?

14. *Induced spectrum.* DNA photolyases convert the energy of light in the near-ultraviolet or visible region (300–500 nm)

into chemical energy to break the cyclobutane ring of pyrimidine dimers. In the absence of substrate, these photoreactivating enzymes do not absorb light of wavelengths longer than 300 nm. Why is the substrate-induced absorption band advantageous?

15. *Missing telomerase.* Cells lacking telomerase can grow for several cell divisions without obvious defects. However, after more cell divisions, such cells tend to show chromosomes that have fused together. Propose an explanation for the formation of the chromosomes.

16. *I need to unwind.* With the assumption that the energy required to break an average base pair in DNA is 10 kJ mol^{-1} (2.4 kcal mol^{-1}), estimate the maximum number of base pairs that could be broken per ATP hydrolyzed by a helicase operating under standard conditions.

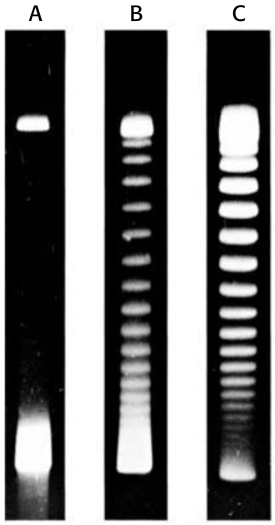
17. *Triplet oxidation.* The oxidation of guanine bases in the context of triplet repeats such as CAGCAGCAG can lead to the expansion of the repeat. Explain.

Mechanism Problem

18. *A revealing analog.* AMP-PNP, the β,γ-imido analog of ATP (p. 833), is hydrolyzed very slowly by most ATPases. The addition of AMP-PNP to topoisomerase II and circular DNA leads to the negative supercoiling of a single molecule of DNA per enzyme. DNA remains bound to the enzyme in the presence of this analog. What does this finding reveal about the catalytic mechanism?

Data Interpretation and Chapter Integration Problems

19. *Like a ladder.* Circular DNA from SV40 virus was isolated and subjected to gel electrophoresis. The results are shown in lane A (the control) of the adjoining gel patterns.



[From W. Keller, *PNAS* 72:2553, 1975.]

(a) Why does the DNA separate in agarose gel electrophoresis? How does the DNA in each band differ?

The DNA was then incubated with topoisomerase I for 5 minutes and again analyzed by gel electrophoresis with the results shown in lane B.

(b) What types of DNA do the various bands represent?

Another sample of DNA was incubated with topoisomerase I for 30 minutes and again analyzed as shown in lane C.

(c) What is the significance of the fact that more of the DNA is in slower-moving forms?

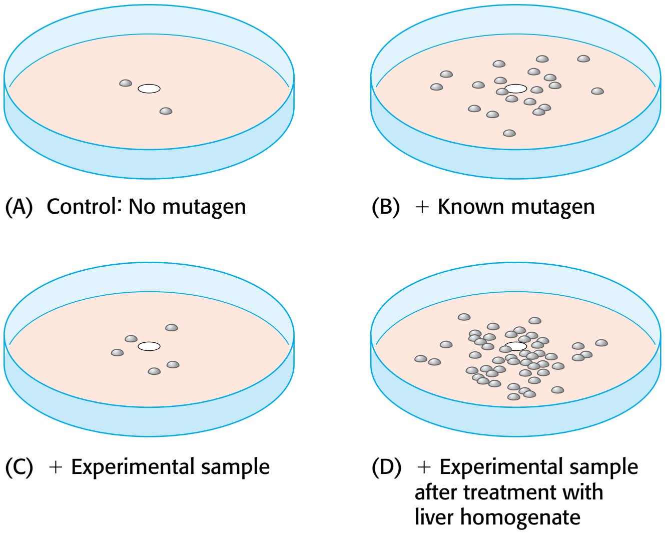
20. *Ames test.* The illustration below shows four petri plates used for the Ames test. A piece of filter paper (white circle in the center of each plate) was soaked in one of four preparations and then placed on a petri plate. The four preparations contained (A) purified water (control), (B) a known mutagen, (C) a chemical whose mutagenicity is under investigation, and (D) the same chemical after treatment with liver homogenate. The number of revertants, visible as colonies on the petri plates, was determined in each case.

(a) What was the purpose of the control plate, which was exposed only to water?

(b) Why was it wise to use a known mutagen in the experimental system?

(c) How would you interpret the results obtained with the experimental compound?

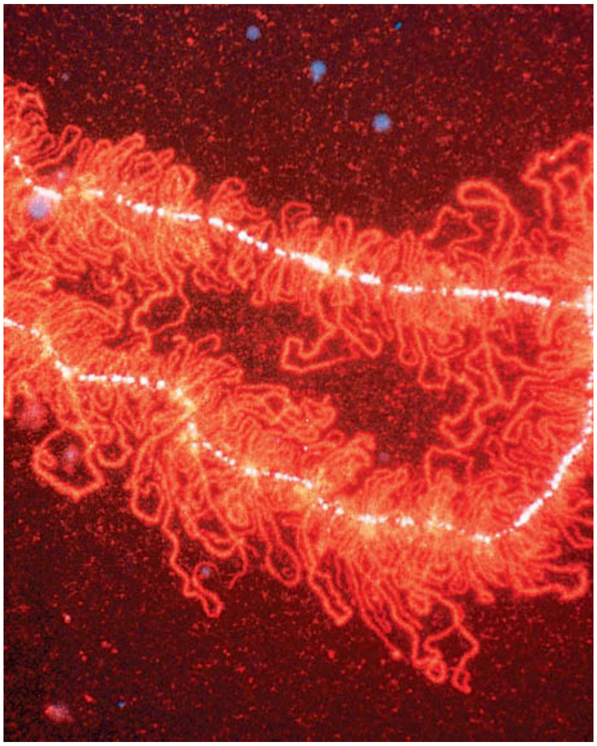
(d) What liver components would you think are responsible for the effects observed in preparation D?

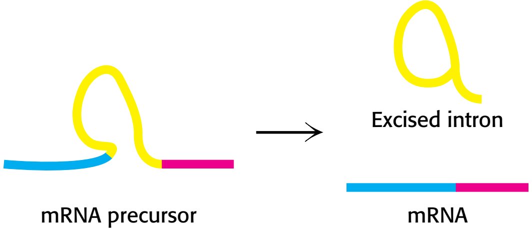

(A) Control: No mutagen
(B) + Known mutagen
(C) + Experimental sample
(D) + Experimental sample after treatment with liver homogenate

CHAPTER

29

RNA Synthesis and Processing





RNA synthesis is a key step in the expression of genetic information. For eukaryotic cells, the initial RNA transcript (the mRNA precursor) is often spliced, removing introns that do not encode protein sequences. Often, the same pre-mRNA is spliced differently in different cell types or at different developmental stages. In the image at the left, proteins associated with RNA splicing (stained with a fluorescent antibody) highlight regions of the newt genome that are being actively transcribed. [(Left) Courtesy of Dr. Mark B. Roth and Dr. Joseph G. Gall.]

DNA stores genetic information in a stable form that can be readily replicated. The expression of this genetic information requires its flow from DNA to RNA and, usually, to protein, as was introduced in Chapter 4. This chapter examines RNA synthesis, or *transcription,* which is the process of synthesizing an RNA transcript with the transfer of sequence information from a DNA template. We begin with a discussion of RNA polymerases, the large and complex enzymes that carry out the synthetic process. We then turn to transcription in bacteria and focus on the three stages of transcription: promoter binding and initiation, elongation of the nascent RNA transcript, and termination. We then examine transcription in eukaryotes, focusing on the distinctions between bacterial and eukaryotic transcription.

RNA transcripts in eukaryotes are extensively modified, as exemplified by the capping of the 5′ end of an mRNA precursor and the addition of a long poly(A) tail to its 3′ end. One of the most striking examples of RNA modification is the splicing of mRNA precursors, which is catalyzed by spliceosomes, protein complexes consisting of small nuclear ribonucleoprotein particles (snRNPs). Remarkably, some RNA molecules can splice themselves in the absence of protein. This landmark discovery by Thomas Cech and Sidney Altman revealed that RNA molecules can serve as catalysts, greatly influencing our view of molecular evolution.

OUTLINE

29.1 RNA Polymerases Catalyze Transcription

29.2 Transcription in Eukaryotes Is Highly Regulated

29.3 The Transcription Products of Eukaryotic Polymerases Are Processed

29.4 The Discovery of Catalytic RNA Was Revealing in Regard to Both Mechanism and Evolution

RNA splicing is not merely a curiosity. Many genetic diseases have been associated with mutations that affect RNA splicing. Moreover, the same pre-mRNA can be spliced differently in various cell types, at different stages of development, or in response to other biological signals. In addition, individual bases in some pre-mRNA molecules are changed in a process called RNA editing. One of the biggest surprises of the sequencing of the human genome was that only about 21,000 genes were identified compared with previous estimates of 100,000 or more. The ability of one gene to encode more than one distinct mRNA by alternative splicing and, hence, more than one protein may play a key role in expanding the repertoire of our genomes.

Furthermore, the investigation of different classes of RNA molecules has been one of the most productive areas of biochemical research in recent years. In the next chapter we will explore the long-known ribosomal and transfer RNA molecules that are central to protein synthesis. We have also encountered microRNAs, molecules for which our understanding is still rapidly expanding. Many other classes of RNAs have been discovered more recently, including long non-coding RNAs, the functions of which are still under active investigation.

RNA synthesis comprises three stages: Initiation, elongation, and termination

RNA synthesis is catalyzed by large enzymes called *RNA polymerases*. The basic biochemistry of RNA synthesis is common to all organisms, commonality that has been beautifully illustrated by the three-dimensional structures of representative RNA polymerases from prokaryotes and eukaryotes (Figure 29.1). Despite substantial differences in size and number of polypeptide subunits, the overall structures of these enzymes are quite similar, revealing a common evolutionary origin.

RNA synthesis, like all biological polymerization reactions, takes place in three stages: *initiation, elongation,* and *termination*. RNA polymerases perform multiple functions in this process:

1. They search DNA for initiation sites, also called *promoter sites* or simply *promoters*. For instance, *E. coli* DNA has about 2000 promoter sites in its 4.8×10^6 bp genome.

2. They unwind a short stretch of double-helical DNA to produce single-stranded DNA templates from which the sequence of bases can be easily read out.

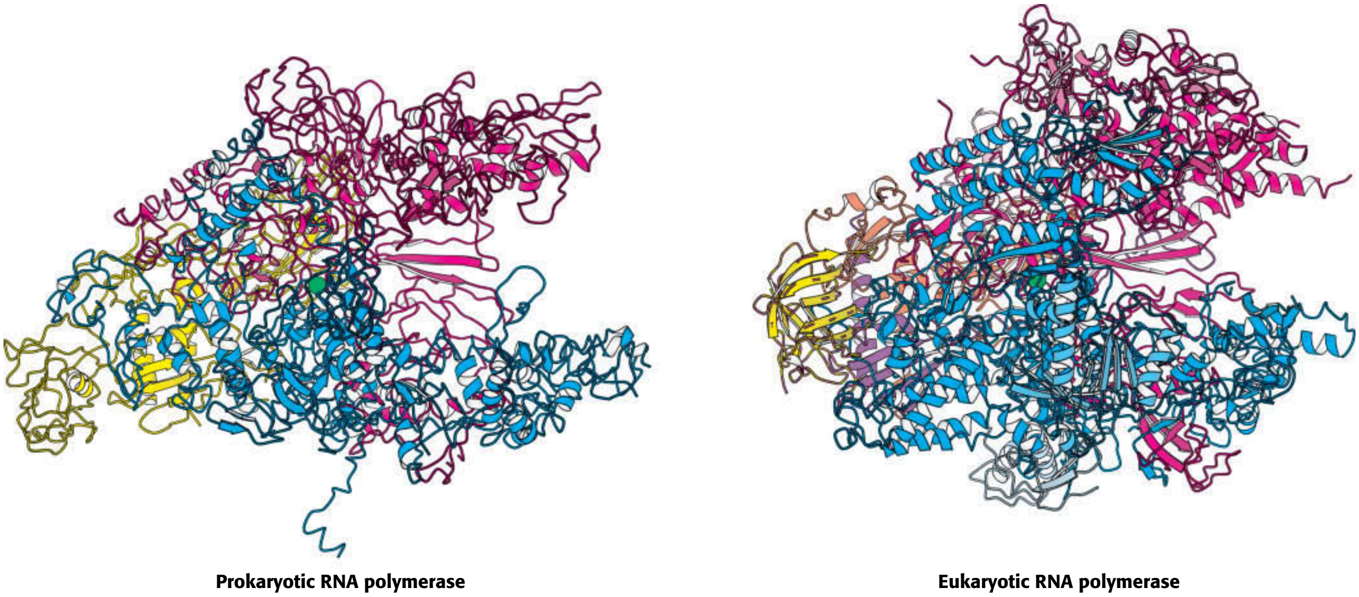

Prokaryotic RNA polymerase **Eukaryotic RNA polymerase**

FIGURE 29.1 RNA polymerase structures. The three-dimensional structures of RNA polymerases from a prokaryote *(Thermus aquaticus)* and a eukaryote *(Saccharomyces cerevisiae).* The two largest subunits for each structure are shown in dark red and dark blue. *Notice* that both structures contain a central metal ion (green) in their active sites, near a large cleft on the right. The similarity of these structures reveals that these enzymes have the same evolutionary origin and have many mechanistic features in common. [Drawn from 1I6V.pdb and 1I6H.pdb.]

3. They select the correct ribonucleoside triphosphate and catalyze the formation of a phosphodiester bond. This process is repeated many times as the enzyme moves along the DNA template. RNA polymerase is completely processive—a transcript is synthesized from start to end by a single RNA polymerase molecule.

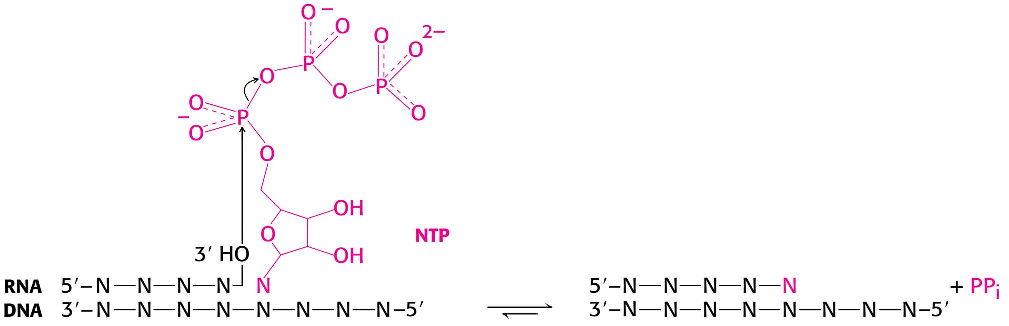
4. They detect termination signals that specify where a transcript ends.

5. They interact with activator and repressor proteins that modulate the rate of transcription initiation over a wide range. Gene expression is controlled substantially at the level of transcription, as will be discussed in detail in Chapters 31 and 32.

The chemistry of RNA synthesis is identical for all forms of RNA, including messenger RNAs, transfer RNAs, ribosomal RNAs, and small regulatory RNAs. The basic steps just outlined apply to all forms. Their synthetic processes differ mainly in regulation, posttranscriptional processing, and the specific RNA polymerase that participates.

29.1 RNA Polymerases Catalyze Transcription

The fundamental reaction of RNA synthesis is the formation of a phosphodiester bond. The 3′-hydroxyl group of the last nucleotide in the chain nucleophilically attacks the α phosphoryl group of the incoming nucleoside triphosphate with the concomitant release of a pyrophosphate.



This reaction is thermodynamically favorable, and the subsequent degradation of the pyrophosphate to orthophosphate locks the reaction in the direction of RNA synthesis. The catalytic sites of RNA polymerases include two metal ions, normally magnesium ions (Figure 29.2). One ion remains tightly bound to the enzyme, whereas the other ion comes in with the nucleoside triphosphate and leaves with the pyrophosphate. Three conserved aspartate residues participate in binding these metal ions.

RNA polymerases are very large, complex enzymes. For example, the core of the RNA polymerase of *E. coli* consists of 5 kinds of subunits with the composition $\alpha_2\beta\beta'\omega$ (Table 29.1). A typical eukaryotic RNA polymerase is larger and more complex, having 12 subunits and a total molecular mass of more than 0.5 million daltons. Despite

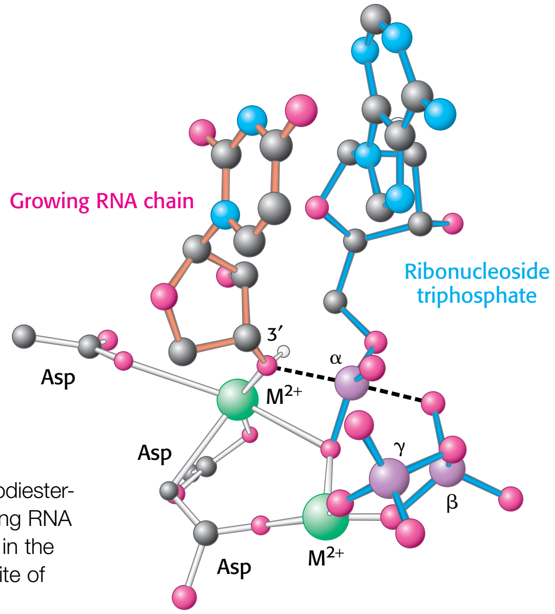


FIGURE 29.2 RNA polymerase active site. A model of the transition state for phosphodiester-bond formation in the active site of RNA polymerase. The 3′-hydroxyl group of the growing RNA chain attacks the α-phosphoryl group of the incoming nucleoside triphosphate, resulting in the release of pyrophosphate. This transition state is structurally similar to that in the active site of DNA polymerase (Figure 28.5).

TABLE 29.1 Subunits of RNA polymerase from *E. coli*

Subunit	Gene	Number	Mass (kDa)
α	*rpoA*	2	37
β	*rpoB*	1	151
β'	*rpoC*	1	155
ω	*rpoZ*	1	10
σ^{70}	*rpoD*	1	70

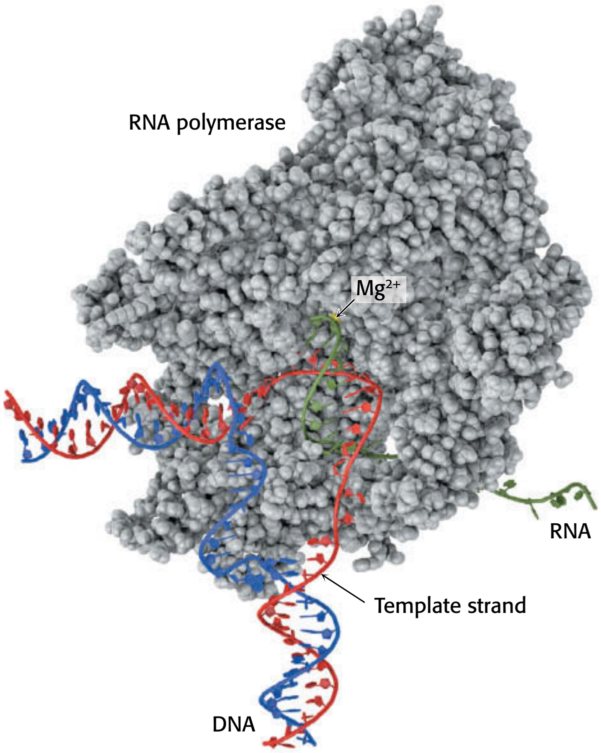


FIGURE 29.3 Transcription bubble. RNA polymerase separates a region of the double helix to form a structure called the "transcription bubble." The red (template) and blue (nontemplate) strands of DNA are shown along with the RNA molecule being synthesized (shown in green). The position of the active site magnesium is indicated. The DNA enters from the left and exits at the bottom.

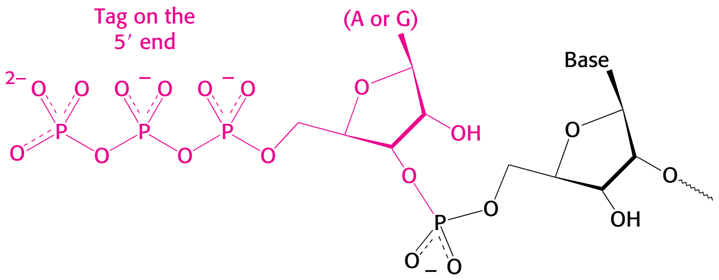
this complexity, the structures of RNA polymerases have been determined in detail by x-ray crystallography in work pioneered by Roger Kornberg and Seth Darst.

The polymerization reactions catalyzed by RNA polymerases take place within a complex in DNA termed a *transcription bubble* (Figure 29.3). This complex consists of double-stranded DNA that has been locally unwound in a region of approximately 17 base pairs. The edges of the bases that normally take part in Watson–Crick base pairs are exposed in the unwound region. We will begin with a detailed examination of the elongation process, including the role of the DNA template read by RNA polymerase and the reactions catalyzed by the polymerase, before returning to the more-complex processes of initiation and termination.

RNA chains are formed de novo and grow in the 5′-to-3′ direction

Let us begin our examination of transcription by considering the DNA template. The first nucleotide (the start site) of a DNA sequence to be transcribed is denoted as +1 and the second one as +2; the nucleotide preceding the start site is denoted as −1. These designations refer to the coding strand of DNA. Recall that the sequence of the *template strand of DNA* is the *complement* of that of the RNA transcript (Figure 29.4). In contrast, the *coding strand of DNA* has the *same* sequence as that of the RNA transcript except for thymine (T) in place of uracil (U). The coding strand is also known as the *sense* (+) *strand,* and the template strand as the *antisense* (−) *strand.*

In contrast with DNA synthesis, *RNA synthesis can start de novo, without the requirement for a primer.* Most newly synthesized RNA chains carry a highly distinctive tag on the 5′ end: the first base at that end is either *pppG* or *pppA.*



The presence of the triphosphate moiety confirms that RNA synthesis starts at the 5′ end.

The dinucleotide shown above is synthesized by RNA polymerase as part of the complex process of initiation, which will be discussed later in the chapter. After initiation takes place, RNA polymerase elongates the nucleic

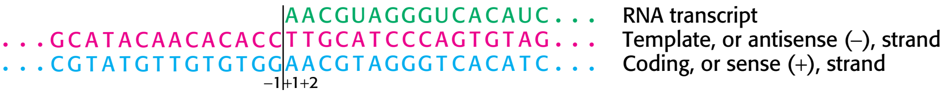

AACGUAGGGUCACAUC . . .
. . . GCATACAACACACCTTGCATCCCAGTGTAG . . .
. . . CGTATGTTGTGTGGAACGTAGGGTCACATC . . .
−1
+1
+2
RNA transcript
Template, or antisense (−), strand
Coding, or sense (+), strand

FIGURE 29.4 Template and coding strands. The template, or antisense (−), strand is complementary in sequence to the RNA transcript.

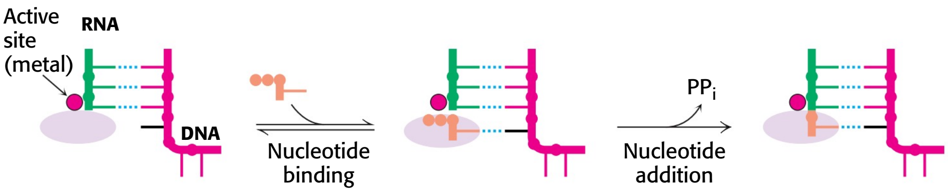


FIGURE 29.5 Elongation mechanism. A ribonucleoside triphosphate binds adjacent to the growing RNA chain and forms a Watson–Crick base pair with a base on the DNA template strand. The 3′-hydroxyl group at the end of the RNA chain attacks the newly bound nucleotide and forms a new phosphodiester bond, releasing pyrophosphate.

acid chain in the following manner (Figure 29.5). A ribonucleoside triphosphate binds in the active site of the RNA polymerase, directly adjacent to the growing RNA chain. The incoming ribonucleoside triphosphate forms a Watson–Crick base pair with the template strand. The 3′-hydroxyl group of the growing RNA chain, oriented and activated by the tightly bound metal ion, attacks the α-phosphoryl group to form a new phosphodiester bond, displacing pyrophosphate.

To proceed to the next step, the RNA–DNA hybrid must move relative to the polymerase to bring the 3′ end of the newly added nucleotide into proper position for the next nucleotide to be added (Figure 29.6). This translocation step does not include breaking any bonds between base pairs and is reversible but, once it has taken place, the addition of the next nucleotide, favored by the triphosphate cleavage and pyrophosphate release and cleavage, drives the polymerization reaction forward.

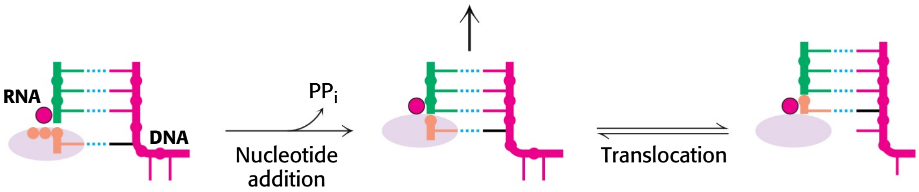


FIGURE 29.6 Translocation. After nucleotide addition, the RNA–DNA hybrid can translocate through the RNA polymerase, bringing a new DNA base into position to base-pair with an incoming nucleoside triphosphate.

The lengths of the RNA–DNA hybrid and of the unwound region of DNA stay rather constant as RNA polymerase moves along the DNA template. The length of the RNA–DNA hybrid is determined by a structure within the enzyme that forces the RNA–DNA hybrid to separate, allowing the RNA chain to exit from the enzyme and the DNA chain to rejoin its DNA partner (Figure 29.7).

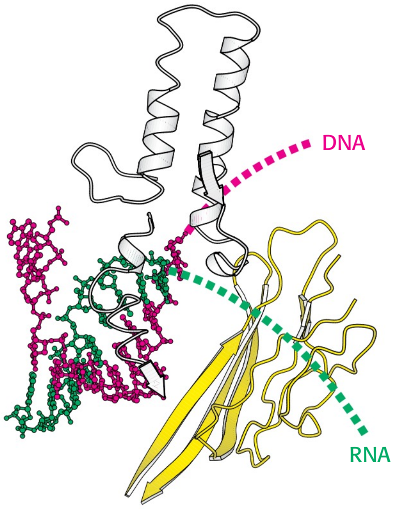


FIGURE 29.7 RNA–DNA hybrid separation. A structure within RNA polymerase forces the separation of the RNA–DNA hybrid. *Notice* that the DNA strand exits in one direction and the RNA product exits in another. [Drawn from 1I6H.pdb.]

RNA polymerases backtrack and correct errors

The RNA–DNA hybrid can also move in the direction opposite that of elongation (Figure 29.8). This backtracking is less favorable energetically than moving forward because it breaks the bonds between a base pair. However, backtracking is very important for *proofreading*. The incorporation of an incorrect nucleotide introduces a non-Watson–Crick base pair. In this case, breaking the bonds between this base pair and backtracking is less costly energetically. After the polymerase has backtracked, the phosphodiester bond one base pair before the one that has just formed is adjacent to the metal ion in the active site. In this position, a hydrolysis reaction in

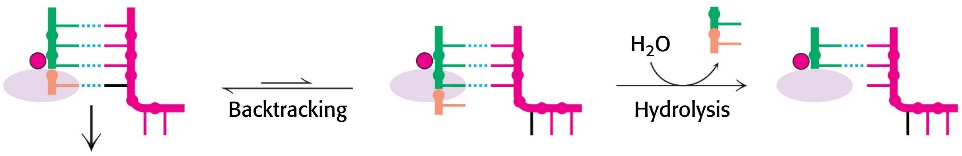


FIGURE 29.8 Backtracking. The RNA–DNA hybrid can occasionally backtrack within the RNA polymerase. In the backtracked position, hydrolysis can take place, producing a configuration equivalent to that after translocation. Backtracking is more likely if a mismatched base is added, facilitating proofreading.

which a water molecule attacks the phosphate can result in the cleavage of the phosphodiester bond and the release of a dinucleotide that includes the incorrect nucleotide.

Studies of single molecules of RNA polymerase have confirmed that the enzymes hesitate and backtrack to correct errors. Furthermore, these proofreading activities are often enhanced by accessory proteins. The final error rate of the order of one mistake per 10^4 or 10^5 nucleotides is higher than that for DNA replication, including all error-correcting mechanisms. The lower fidelity of RNA synthesis can be tolerated because mistakes are not transmitted to progeny. For most genes, many RNA transcripts are synthesized; a few defective transcripts are unlikely to be harmful.

RNA polymerase binds to promoter sites on the DNA template to initiate transcription

The elongation process is common to all organisms. In contrast, the processes of initiation and termination differ substantially in bacteria and eukaryotes. We begin with a discussion of these processes in bacteria, starting with initiation of transcription. The bacterial RNA polymerase discussed earlier with the composition $\alpha_2\beta\beta'\omega$ is referred to as the *core enzyme*. The inclusion of an additional subunit produces the *holoenzyme* with composition $\alpha_2\beta\beta'\omega\sigma$. The σ subunit helps find sites on DNA where transcription begins, referred to as *promoter sites* or, simply, *promoters*. At these sites, the σ subunit participates in the initiation of RNA synthesis and then dissociates from the rest of the enzyme.

Transcription starts here
↓
5′ −10 3′
(A) GATATAATGAGCAAA
(B) CGTATAATTTGACCA
(C) GTTAACTAGTACGCA
(D) GTGATACTGAGCACA
(E) GTTTTCATGCCTCCA
TATAAT

FIGURE 29.9 Bacterial promoter sequences. A comparison of five sequences from prokaryotic promoters reveals a recurring sequence of TATAAT centered on position −10. The −10 consensus sequence (in red) was deduced from a large number of promoter sequences. The sequences are from the (A) *uvrD*, (B) *uncI*, and (C) *trp* operons of *E. coli,* from (D) λ phage, and from (E) ϕX174 phage.

Sequences upstream of the promoter site are important in determining where transcription begins. A striking pattern was evident when the sequences of bacterial promoters were compared. *Two common motifs are present on the upstream side of the transcription start site*. They are known as the *−10 sequence* and the *−35 sequence* because they are centered at about 10 and 35 nucleotides upstream of the start site. The region containing these sequences is called the *core promoter*. The −10 and −35 sequences are each 6 bp long. Their *consensus sequences*, deduced from analyses of many promoters (Figure 29.9), are

−35 −10 +1
5′~~TTGACA~~~~~~TATAAT~~~ Start site

Promoters differ markedly in their efficacy. Some genes are transcribed frequently—as often as every 2 seconds in *E. coli*. The promoters for these genes are referred to as *strong promoters*. In contrast, other genes are transcribed much less frequently, about once in 10 minutes; the promoters for these genes are *weak promoters*. The −10 and −35 regions of most strong promoters have sequences that correspond closely to the consensus sequences, whereas weak promoters tend to have multiple substitutions at these sites. Indeed, mutation of a single base in either the −10 sequence or the −35 sequence can diminish promoter activity. The distance between these conserved sequences also is important; a separation of 17 nucleotides is optimal. Thus, *the efficiency or strength of a promoter sequence serves to regulate transcription*. Regulatory proteins that bind to specific sequences near promoter sites and interact with RNA polymerase (Chapter 31) also markedly influence the frequency of transcription of many genes.

Outside the core promoter in a subset of highly expressed genes is the *upstream element* (also called the UP element for *up*stream element). This sequence is present from 40 to 60 nucleotides upstream of the transcription start site. The UP element is bound by the α subunit of RNA polymerase and serves to increase the efficiency of transcription by creating an additional interaction site for the polymerase.

Sigma subunits of RNA polymerase recognize promoter sites

To initiate transcription, the $\alpha_2\beta\beta'\omega$ core of RNA polymerase must bind the promoter. However, it is the σ *subunit* that makes this binding possible by enabling *RNA polymerase to recognize promoter sites*. In the presence of the σ subunit, the RNA polymerase binds weakly to the DNA and slides along the double helix until it dissociates or encounters a promoter. The σ subunit recognizes the promoter through several interactions with the nucleotide bases of the promoter DNA. The structure of a bacterial RNA polymerase holoenzyme bound to a promoter site shows the σ subunit interacting with DNA at the −10 and −35 regions essential to promoter recognition (Figure 29.10). Therefore, *the σ subunit is responsible for the specific binding of the RNA polymerase to a promoter site on the template DNA*. The σ subunit is generally released when the nascent RNA chain reaches 9 or 10 nucleotides in length. After its release, it can assist initiation by another core enzyme. Thus, the σ *subunit acts catalytically*.

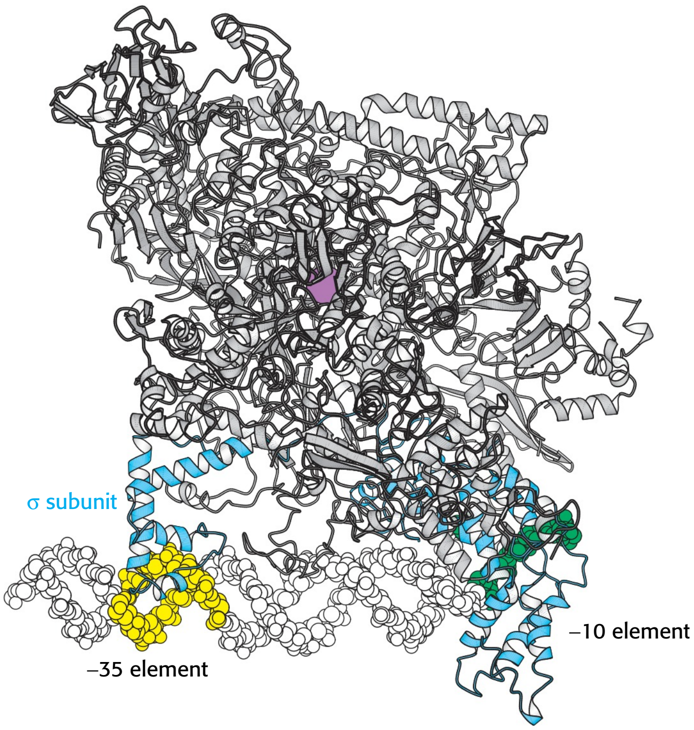


FIGURE 29.10 RNA polymerase holoenzyme complex. *Notice* that the σ subunit (blue) of the bacterial RNA polymerase holoenzyme makes sequence-specific contacts with the −10 (green) and −35 promoter elements (yellow). The active site of the polymerase is revealed by bound metal ion (purple). [Drawn from 1L9Z.pdb].

E. coli has seven distinct σ factors for recognizing several types of promoter sequences in *E. coli* DNA. The type that recognizes the consensus sequences described earlier is called σ^{70} because it has a mass of 70 kDa. A different σ factor comes into play when the temperature is raised abruptly. *E. coli* responds by synthesizing σ^{32}, which recognizes the promoters of *heat-shock genes*. These promoters exhibit −10 sequences that are somewhat different from the −10 sequence for standard promoters (Figure 29.11). The increased transcription of heat-shock genes leads to the coordinated synthesis of a series of protective proteins. Other σ factors respond to environmental conditions, such as nitrogen starvation. These findings demonstrate that *σ plays the key role in determining where RNA polymerase initiates transcription.*

Some other bacteria contain a much larger number of σ factors. For example, the genome of the soil bacterium *Streptomyces coelicolor* encodes more than 60 σ factors recognized on the basis of their amino acid sequences. This repertoire allows these cells to adjust their gene-expression programs to the wide range of conditions, in regard to nutrients and competing organisms, that they may experience.

RNA polymerases must unwind the template double helix for transcription to take place

Although RNA polymerases can search for promoter sites when bound to double-helical DNA, a segment of the DNA double helix must be unwound before synthesis can begin. The transition from the *closed promoter complex*

	−35		−10		
5′~~~	TTGACA	~~~~~~~~	TATAAT	~~~~3′	Standard promoter
5′~~~	TNNCNCNCTTGAA	~~~~~~~~	CCCATNT	~~~~3′	Heat-shock promoter
5′~~~	CTGGGNA	~~~~~~~~	TTGCA	~~~~3′	Nitrogen-starvation promoter

FIGURE 29.11 Alternative promoter sequences. A comparison of the consensus sequences of standard, heat-shock, and nitrogen-starvation promoters of *E. coli.* These promoters are recognized by σ^{70}, σ^{32}, and σ^{54}, respectively.

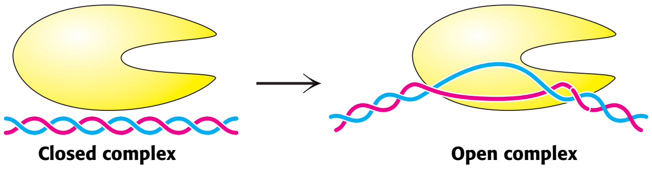


FIGURE 29.12 DNA unwinding. The transition from the closed promoter complex to the open promoter complex requires the unwinding of approximately 17 base pairs of DNA.

(in which DNA is double helical) to the *open promoter complex* (in which a DNA segment is unwound) is an essential event in transcription (Figure 29.12). The free energy necessary to break the bonds between approximately 17 base pairs in the double helix is derived from additional interactions that are possible when the DNA distorts to wrap around the RNA polymerase and from interactions between the single-stranded DNA regions and other parts of the enzyme. These interactions stabilize the open promoter complex and help pull the template strand into the active site. The −35 element remains in a double-helical state, whereas the −10 element is unwound. The stage is now set for the formation of the first phosphodiester bond of the new RNA chain.

Elongation takes place at transcription bubbles that move along the DNA template

The elongation phase of RNA synthesis begins with the formation of the first phosphodiester bond. Repeated cycles of nucleotide addition can take place at this point. However, until about 10 nucleotides have been added, RNA polymerase sometimes releases the short RNA, which dissociates from the DNA. Once RNA polymerase passes this point, the enzyme stays bound to its template until a termination signal is reached. The region containing RNA polymerase, DNA, and nascent RNA corresponds to the transcription bubble (Figure 29.13). The newly synthesized RNA forms a hybrid helix with the template DNA strand. This RNA–DNA helix is about 8 bp long, which corresponds to nearly one turn of a double helix. The 3′-hydroxyl group of the RNA in this hybrid helix is positioned so that it can attack the α-phosphorus atom of an incoming ribonucleoside triphosphate. The core enzyme also contains a binding site for the coding strand of DNA. About 17 bp of DNA are unwound throughout the elongation phase, as in the initiation phase. The transcription bubble moves a distance of 170 Å (17 nm) in a second, which corresponds to a rate of elongation of about 50 nucleotides per second. Although rapid, it is much slower than the rate of DNA synthesis, which is 800 nucleotides per second.

Sequences within the newly transcribed RNA signal termination

In bacteria, the termination of transcription is as precisely controlled as its initiation. In the termination phase of transcription, the formation of phosphodiester bonds ceases, the RNA–DNA hybrid dissociates, the unwound region of DNA rewinds, and RNA polymerase releases the DNA. What determines where transcription is terminated? *The transcribed regions of DNA*

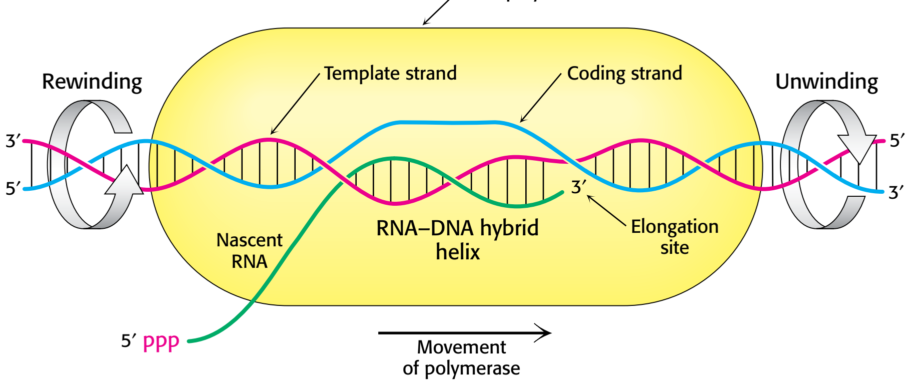


FIGURE 29.13 Transcription bubble. A schematic representation of a transcription bubble in the elongation of an RNA transcript. Duplex DNA is unwound at the forward end of RNA polymerase and rewound at its rear end. The RNA–DNA hybrid rotates during elongation.

templates contain stop signals. The simplest one is a *palindromic GC-rich region followed by an AT-rich region*. The RNA transcript of this DNA palindrome is self-complementary (Figure 29.14). Hence, its bases can pair to form a hairpin structure with a stem and loop, a structure favored by its high content of G and C residues. Guanine–cytosine base pairs are more stable than adenine–thymine pairs because of the extra hydrogen bond in the base pair. This stable hairpin is followed by a sequence of four or more uracil residues, which also are crucial for termination. The RNA transcript ends within or just after them.

How does this combination hairpin–oligo(U) structure terminate transcription? First, RNA polymerase likely pauses immediately after it has synthesized a stretch of RNA that folds into a hairpin. Furthermore, the RNA–DNA hybrid helix produced after the hairpin is unstable because its rU–dA base pairs are the weakest of the four kinds. Hence, the pause in transcription caused by the hairpin permits the weakly bound *nascent RNA to dissociate from the DNA template and then from the enzyme*. The solitary DNA template strand rejoins its partner to re-form the DNA duplex, and the transcription bubble closes.

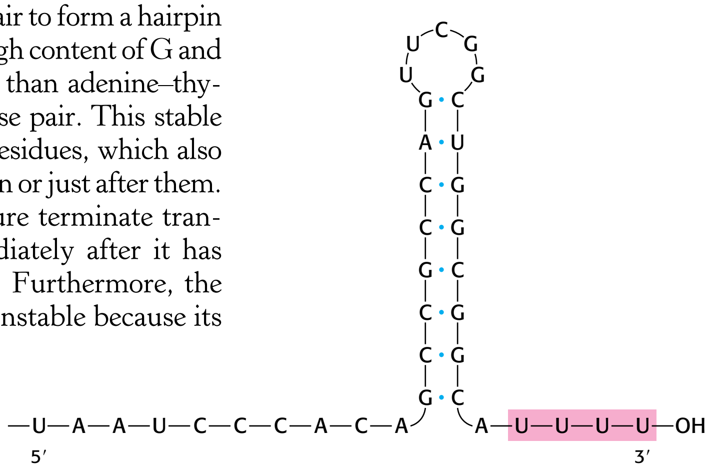
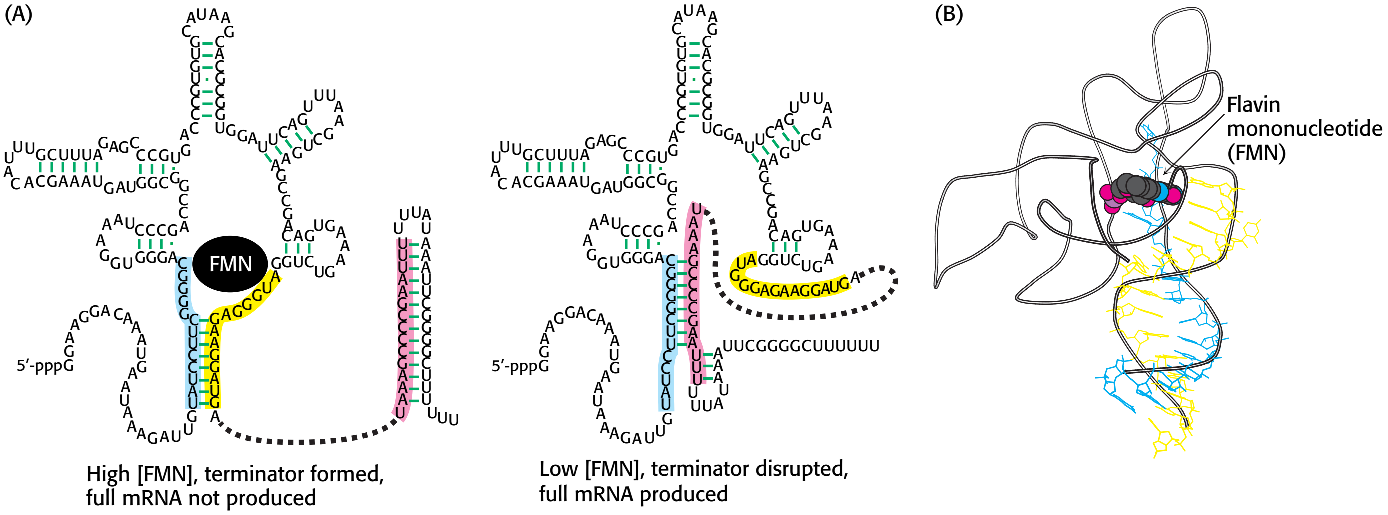


FIGURE 29.14 Termination signal. A termination signal found at the 3′ end of an mRNA transcript consists of a series of bases that form a stable stem-loop structure and a series of U residues.

Some messenger RNAs directly sense metabolite concentrations

As we shall explore in Chapters 31 and 32, the expression of many genes is controlled in response to the concentrations of metabolites and signaling molecules within cells. One set of control mechanisms depends on the remarkable ability of some mRNA molecules to form special secondary structures, some of which are capable of directly binding small molecules. These structures are termed *riboswitches*. Consider a riboswitch that controls the synthesis of genes that participate in the biosynthesis of riboflavin in *Bacillus subtilis* (Figure 29.15). When flavin mononucleotide (FMN), a key intermediate in riboflavin biosynthesis, is present at high concentration, it binds to a conformation of the RNA transcript with an FMN-binding pocket that also includes a hairpin structure that favors premature termination. By trapping the RNA transcript in this termination-favoring conformation, FMN prevents the production of functional mRNA. However, when FMN is present at low concentration, it does not readily bind to RNA, and an alternative conformation forms without the terminator, allowing the production of the full-length mRNA. The occurrence of riboswitches serves as a vivid illustration of how RNAs are capable of forming

FIGURE 29.15 Riboswitch. (A) The 5′-end of an mRNA that encodes proteins engaged in the production of flavin mononucleotide (FMN) folds to form a structure that is stabilized by binding FMN. This structure includes a terminator that leads to premature termination of the mRNA. At lower concentrations of FMN, an alternative structure that lacks the terminator is formed, leading to the production of full-length mRNA. (B) The three-dimensional structure of a related FMN-binding riboswitch bound to FMN. The blue and yellow stretches correspond to regions highlighted in the same colors in part A. *Notice* how the yellow strand contacts the bound FMN, stabilizing the structure. [Drawn from 3F2Q.pdb].



High [FMN], terminator formed, full mRNA not produced

Low [FMN], terminator disrupted, full mRNA produced

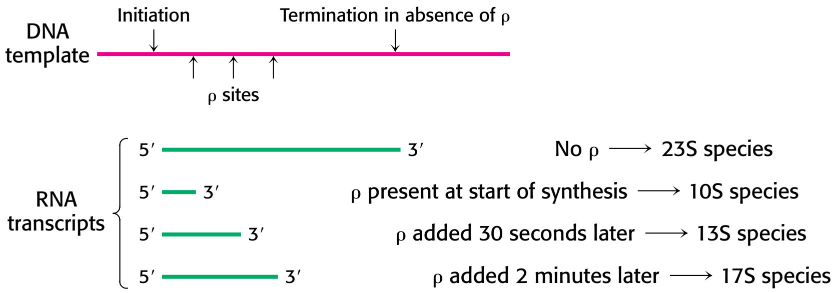

DNA template
Initiation
Termination in absence of ρ
ρ sites
RNA transcripts
5′ 3′ No ρ ⟶ 23S species
5′ 3′ ρ present at start of synthesis ⟶ 10S species
5′ 3′ ρ added 30 seconds later ⟶ 13S species
5′ 3′ ρ added 2 minutes later ⟶ 17S species

FIGURE 29.16 Effect of ρ protein on the size of RNA transcripts.

elaborate, functional structures, though we tend to depict them as simple lines in the absence of specific information.

The *rho* protein helps to terminate the transcription of some genes

RNA polymerase needs no help to terminate transcription at a hairpin followed by several U residues. At other sites, however, termination requires the participation of an additional factor. This discovery was prompted by the observation that some RNA molecules synthesized in vitro by RNA polymerase acting alone are *longer* than those made in vivo. The missing factor, a protein that caused the correct termination, was isolated and named *rho* (ρ). Additional information about the action of ρ was obtained by adding this termination factor to an incubation mixture at various times after the initiation of RNA synthesis (Figure 29.16). RNAs with sedimentation coefficients of 10S, 13S, and 17S were obtained when ρ was added at initiation, a few seconds after initiation, and 2 minutes after initiation, respectively. If no ρ was added, transcription yielded a 23S RNA product. It is evident that the template contains at least three termination sites that respond to ρ (yielding 10S, 13S, and 17S RNA) and one termination site that does not (yielding 23S RNA). Thus, specific termination at a site producing 23S RNA can take place in the absence of ρ. However, ρ detects additional termination signals that are not recognized by RNA polymerase alone.

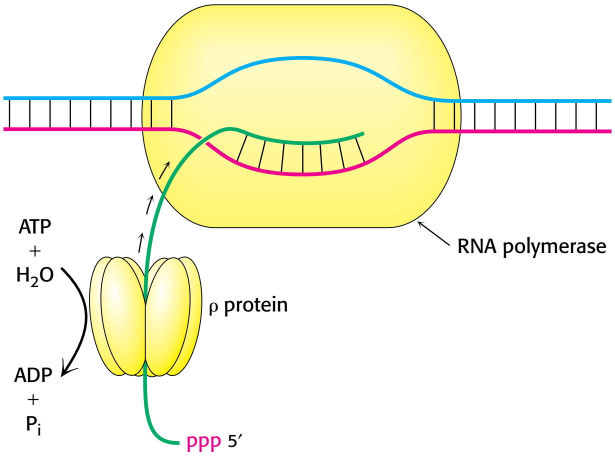


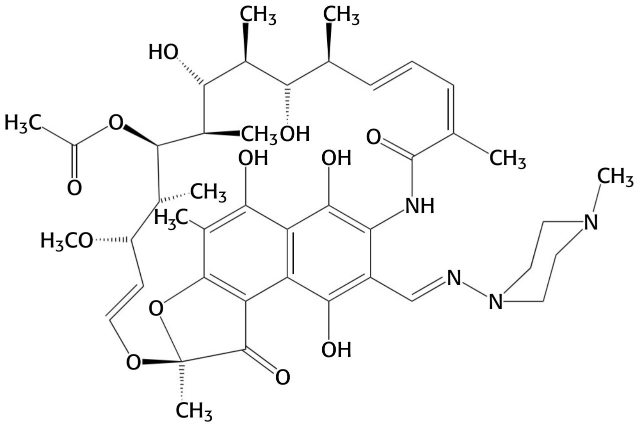
FIGURE 29.17 Mechanism for the termination of transcription by ρ protein. This protein is an ATP-dependent helicase that binds the nascent RNA chain and pulls it away from RNA polymerase and the DNA template.

How does ρ provoke the termination of RNA synthesis? *A key clue is the finding that ρ hydrolyzes ATP in the presence of single-stranded RNA but not in the presence of DNA or duplex RNA.* Hexameric ρ is a helicase, homologous to the helicases that we encountered in our consideration of DNA replication (Section 28.1). A stretch of nucleotides is bound in such a way that the RNA passes through the center of the structure (Figure 29.17). The ρ protein is brought into action by sequences located in the nascent RNA that are rich in cytosine and poor in guanine. The helicase activity of ρ enables the protein to pull the nascent RNA while pursuing RNA polymerase. When ρ catches RNA polymerase at the transcription bubble, it breaks the RNA–DNA hybrid helix by functioning as an RNA–DNA helicase.

Proteins in addition to ρ may provoke termination. For example, the *nusA protein* enables RNA polymerase in *E. coli* to recognize a characteristic class of termination sites. *A common feature of protein-independent and protein-dependent termination is that the functioning signals lie in newly synthesized RNA rather than in the DNA template.*

Some antibiotics inhibit transcription

Many antibiotics are highly specific inhibitors of biological processes in bacteria. Rifampicin and actinomycin are two antibiotics that inhibit bacterial transcription, although in quite different ways. *Rifampicin* is a semisynthetic derivative of *rifamycins,* which are derived from a strain of *Amycolatopsis* that is related to the bacterium that causes strep throat.



Rifampicin

This antibiotic *specifically inhibits the initiation of RNA synthesis*. Rifampicin interferes with the formation of the first few phosphodiester bonds in the RNA chain. The structure of a complex between a prokaryotic RNA polymerase and rifampicin reveals that the antibiotic blocks the channel into which the RNA–DNA hybrid generated by the enzyme must pass (Figure 29.18). The binding site is 12 Å from the active site itself. Rifampicin can inhibit only the initiation of transcription, not elongation, because the RNA–DNA hybrid present in the enzyme during elongation prevents the antibiotic from binding. The pocket in which rifampicin binds is conserved among bacterial RNA polymerases, but not eukaryotic polymerases, and so rifampicin can be used as an antibiotic in antituberculosis therapy.

Actinomycin D, a peptide-containing antibiotic from a different strain of *Streptomyces,* inhibits transcription by an entirely different mechanism. *Actinomycin D binds tightly and specifically to double-helical DNA and thereby prevents it from being an effective template for RNA synthesis.* The results of spectroscopic, hydrodynamic, and structural studies of complexes of actinomycin D and DNA reveal that the phenoxazone ring of actinomycin slips in

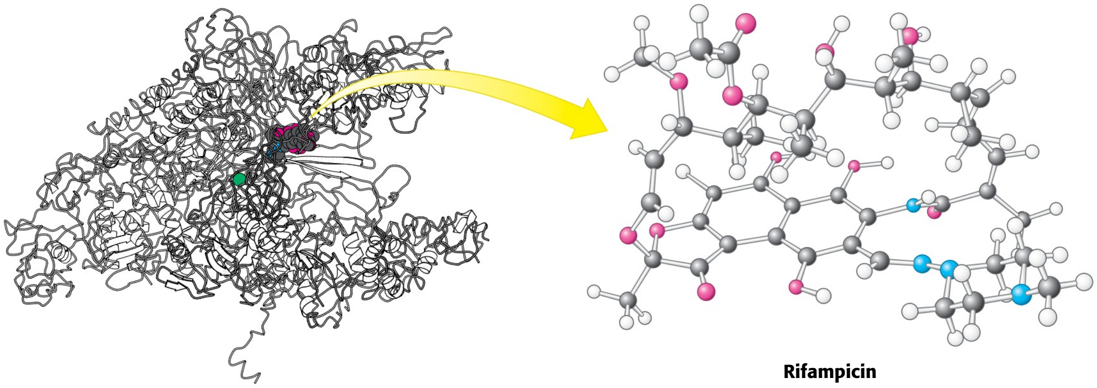


FIGURE 29.18 Antibiotic action. Rifampicin binds to a pocket in the channel that is normally occupied by the newly formed RNA–DNA hybrid. Thus, the antibiotic blocks elongation after only two or three nucleotides have been added.

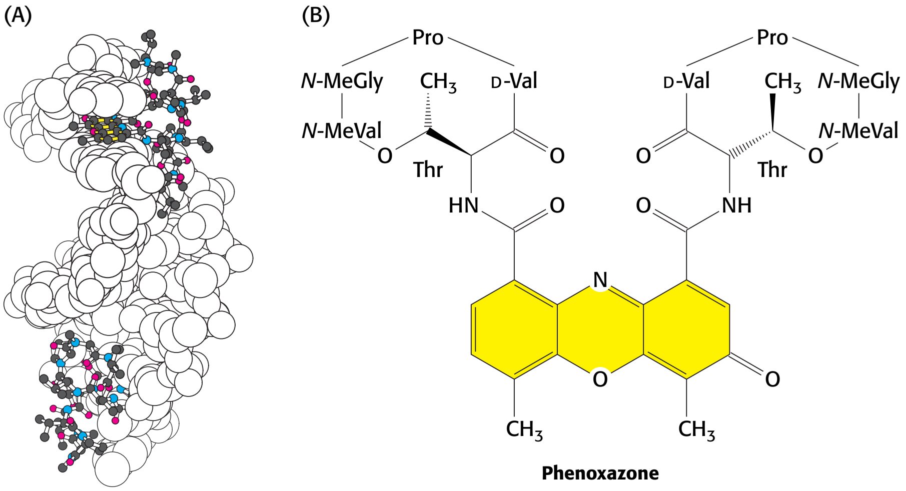


FIGURE 29.19 Actinomycin–DNA complex structure. (A) The structure of a complex between a DNA duplex (shown as a space-filling model) and actinomycin B (shown as a ball-and-stick model). Two actinomycin B molecules are bound in the complex. (B) The structure of actinomycin B showing the phenoxazone ring. *Notice* how the phenoxazone (yellow) slides between base pairs of the DNA. Abbreviation: Me, methyl. [Drawn from 1I3W.pdb]

between base pairs in DNA (Figure 29.19). This mode of binding is called *intercalation*. At low concentrations, actinomycin D inhibits transcription without significantly affecting DNA replication or protein synthesis. Hence, *actinomycin D is extensively used as a highly specific inhibitor of the formation of new RNA in both prokaryotic and eukaryotic cells.* Its ability to inhibit the growth of rapidly dividing cells makes it an effective therapeutic agent in the treatment of some cancers.

Precursors of transfer and ribosomal RNA are cleaved and chemically modified after transcription in prokaryotes

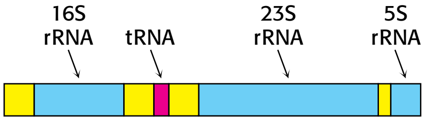
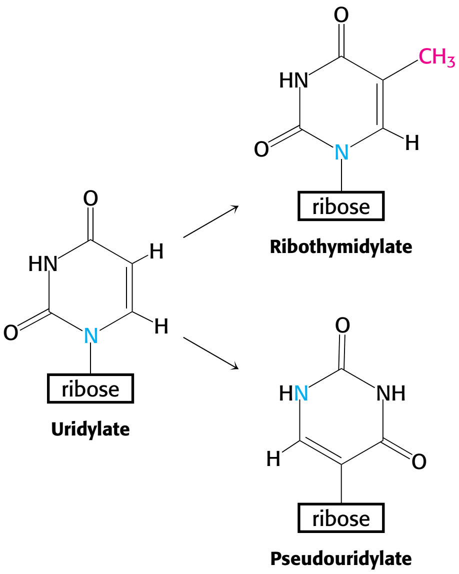


FIGURE 29.20 Primary transcript. Cleavage of this transcript produces 5S, 16S, and 23S rRNA molecules and a tRNA molecule. Spacer regions are shown in yellow.

In prokaryotes, messenger RNA molecules undergo little or no modification after synthesis by RNA polymerase. Indeed, many mRNA molecules are translated while they are being transcribed. In contrast, *transfer RNA (tRNA) and ribosomal RNA (rRNA) molecules are generated by cleavage and other modifications of nascent RNA chains*. For example, in *E. coli,* the three rRNAs and a tRNA are excised from a single primary RNA transcript that also contains spacer regions (Figure 29.20). Other transcripts contain arrays of several kinds of tRNA or of several copies of the same tRNA. The nucleases that cleave and trim these precursors of rRNA and tRNA are highly precise. *Ribonuclease P* (RNase P), for example, generates the correct 5′ terminus of all tRNA molecules in *E. coli.* Sidney Altman and his coworkers showed that this interesting enzyme contains a catalytically active RNA molecule. *Ribonuclease III* (RNase III) excises 5S, 16S, and 23S rRNA precursors from the primary transcript by cleaving double-helical hairpin regions at specific sites.

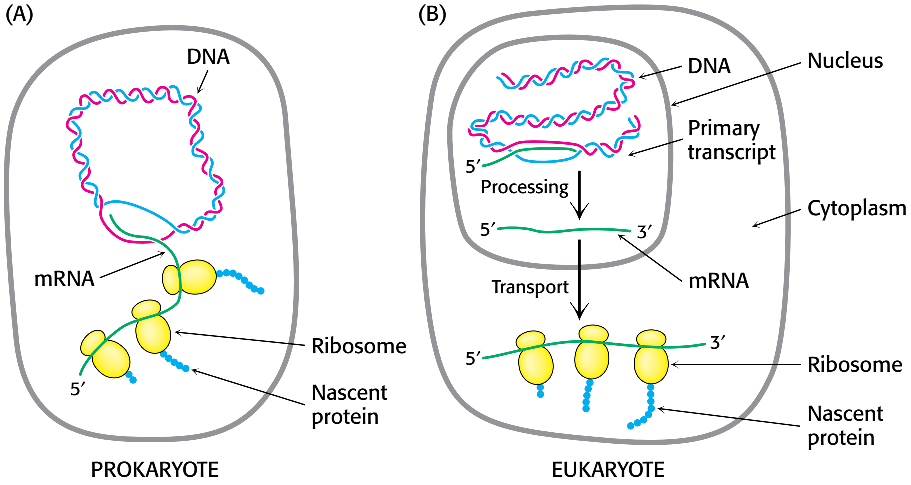
A second type of processing is the *addition of nucleotides to the termini of some RNA chains*. For example, CCA, a terminal sequence required for the function of all tRNAs, is added to the 3′ ends of tRNA molecules for which this terminal sequence is not encoded in the DNA. The enzyme that catalyzes the addition of CCA is atypical for an RNA polymerase in that it does not use a DNA template. A third type of processing is the *modification of bases and ribose units* of ribosomal RNAs. In prokaryotes, some bases of rRNA are methylated. Unusual bases are found in all tRNA molecules. They are formed by the enzymatic modification of a standard ribonucleotide in a tRNA precursor. For example, uridylate residues are modified after transcription to form *ribothymidylate* and *pseudouridylate*. These modifications generate diversity, allowing greater structural and functional versatility.


Uridylate
Ribothymidylate
Pseudouridylate
ribose
CH3
HN
NH
N
O
H

29.2 Transcription in Eukaryotes Is Highly Regulated

We turn now to transcription in eukaryotes, a much more complex process than in bacteria. Eukaryotic cells have a remarkable ability to regulate precisely the time at which each gene is transcribed and how much RNA is produced. This ability has allowed some eukaryotes to evolve into multicellular organisms, with distinct tissues. *That is, multicellular eukaryotes use differential transcriptional regulation to create different cell types*. Gene expression is influenced by three important characteristics unique to eukaryotes: the nuclear membrane, complex transcriptional regulation, and RNA processing.

1. *The Nuclear Membrane. In eukaryotes, transcription and translation take place in different cellular compartments:* transcription takes place in the membrane-bounded nucleus, whereas translation takes place outside the nucleus in the cytoplasm. In bacteria, the two processes are closely coupled (Figure 29.21). Indeed, the translation of bacterial mRNA begins while


(A)
DNA
mRNA
5′
Ribosome
Nascent protein
(B)
DNA
Nucleus
Primary transcript
5′
Processing
5′
3′
Cytoplasm
mRNA
Transport
5′
3′
Ribosome
Nascent protein

PROKARYOTE

EUKARYOTE

FIGURE 29.21 Transcription and translation. These two processes are closely coupled in prokaryotes whereas they are spatially and temporally separate in eukaryotes. (A) In prokaryotes, the primary transcript serves as mRNA and is used immediately as the template for protein synthesis. (B) In eukaryotes, mRNA precursors are processed and spliced in the nucleus before being transported to the cytoplasm for translation into protein. [Information from J. Darnell, H. Lodish, and D. Baltimore. *Molecular Cell Biology*, 2d ed. (Scientific American Books, 1990), p. 230.]

the transcript is still being synthesized. *The spatial and temporal separation of transcription and translation enables eukaryotes to regulate gene expression in much more intricate ways, contributing to the richness of eukaryotic form and function.*

2. *Complex Transcriptional Regulation*. Like bacteria, eukaryotes rely on conserved sequences in DNA to regulate the initiation of transcription. But bacteria have only three promoter elements (the −10, −35, and UP elements), whereas eukaryotes use a variety of types of promoter elements, each identified by its own conserved sequence. Not all possible types will be present together in the same promoter. *In eukaryotes, elements that regulate transcription can be found at a variety of locations in DNA,* upstream or downstream of the start site and sometimes at distances much farther from the start site than in prokaryotes. For example, enhancer elements located on DNA far from the start site increase the promoter activity of specific genes.

3. *RNA Processing*. Although both bacteria and eukaryotes modify RNA, *eukaryotes very extensively process nascent RNA destined to become mRNA.* This processing includes modifications to both ends and, most significantly, splicing out segments of the primary transcript. RNA processing is described in Section 29.3.

Three types of RNA polymerase synthesize RNA in eukaryotic cells

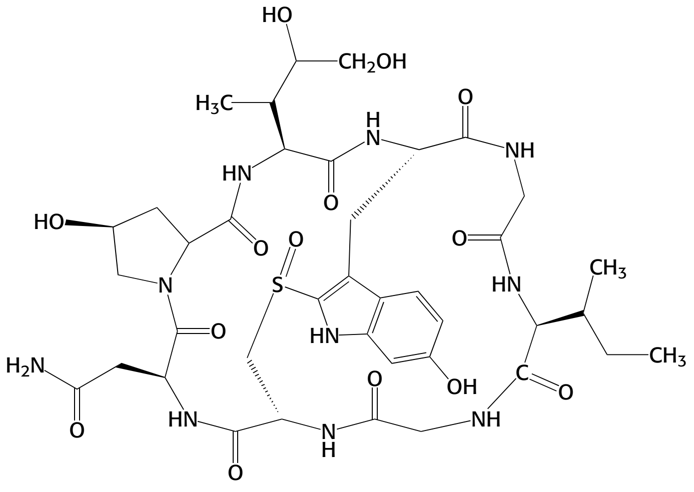
In bacteria, RNA is synthesized by a single kind of polymerase. In contrast, the nucleus of a typical eukaryotic cell contains three types of RNA polymerase differing in template specificity and location in the nucleus (Table 29.2). All these polymerases are large proteins, containing from 8 to 14 subunits and having total molecular masses greater than 500 kDa. *RNA polymerase I* is located in specialized structures within the nucleus called nucleoli, where it transcribes the tandem array of genes for 18S, 5.8S, and 28S rRNA. The other rRNA molecule (5S rRNA) and all the tRNA molecules are synthesized by *RNA polymerase III,* which is located in the nucleoplasm rather than in nucleoli. *RNA polymerase II,* which also is located in the nucleoplasm, synthesizes the precursors of mRNA as well as several small RNA molecules, such as those of the splicing apparatus and many of the precursors to small regulatory RNAs.

Although all eukaryotic RNA polymerases are homologous to one another and to prokaryotic RNA polymerases, RNA polymerase II contains a unique *carboxyl-terminal domain* on the 220-kDa subunit called the CTD; this domain is unusual because it contains multiple repeats of a YSPTSPS consensus sequence. The activity of RNA polymerase II is regulated by phosphorylation mainly on the serine residues of the CTD.

Another major distinction among the polymerases lies in their responses to the toxin *α-amanitin,* a cyclic octapeptide that contains several modified amino acids.

TABLE 29.2 Eukaryotic RNA polymerases

Type	Location	Cellular transcripts	Effects of α-amanitin
I	Nucleolus	18S, 5.8S, and 28S rRNA	Insensitive
II	Nucleoplasm	mRNA precursors and snRNA	Strongly inhibited
III	Nucleoplasm	tRNA and 5S rRNA	Inhibited by high concentrations



α-Amanitin

***Amanita phalloides*, also called the *death cap*.**
[Jacana/Science Source.]

α-Amanitin is produced by the poisonous mushroom *Amanita phalloides,* which is also called the *death cap* or the *destroying angel*. More than a hundred deaths result worldwide each year from the ingestion of poisonous mushrooms. α-Amanitin binds very tightly (K_d = 10 nM) to RNA polymerase II and thereby blocks the elongation phase of RNA synthesis. Higher concentrations of α-amanitin (1 μM) inhibit polymerase III, whereas polymerase I is insensitive to this toxin. This pattern of sensitivity is highly conserved throughout the animal and plant kingdoms.

Eukaryotic polymerases also differ from each other in the promoters to which they bind. Eukaryotic genes, like prokaryotic genes, require promoters for transcription initiation. Like prokaryotic promoters, eukaryotic promoters consist of conserved sequences that attract the polymerase to the start site. However, eukaryotic promoters differ distinctly in sequence and position, depending on the type of RNA polymerase to which they bind (Figure 29.22).

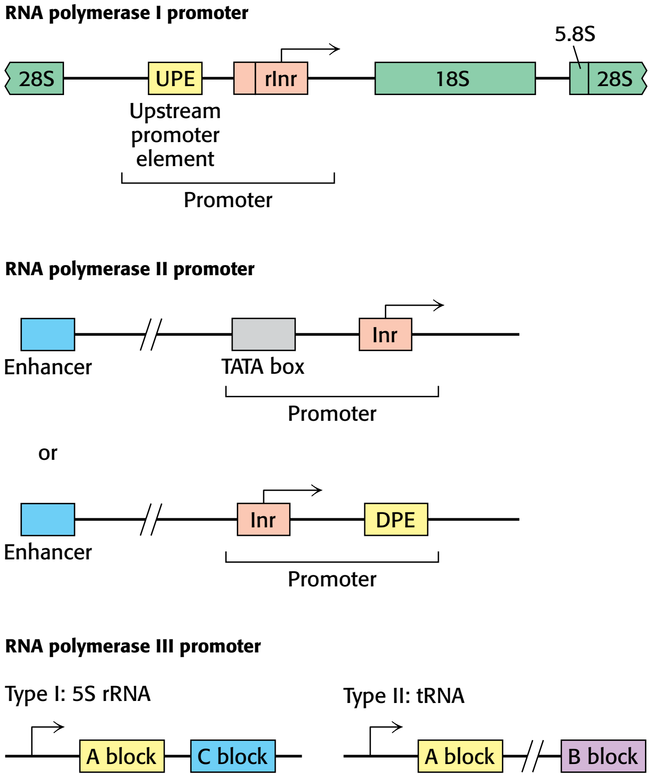


FIGURE 29.22 Common eukaryotic promoter elements. Each eukaryotic RNA polymerase recognizes a set of promoter elements—sequences in DNA that promote transcription. The RNA polymerase I promoter consists of a ribosomal initiator (rInr) and an upstream promoter element (UPE). The RNA polymerase II promoter likewise includes an initiator element (Inr) and may also include either a TATA box or a downstream promoter element (DPE). Separate from the promoter region, enhancer elements bind specific transcription factors. RNA polymerase III promoters consist of conserved sequences that lie within the transcribed genes.

1. *RNA Polymerase I.* The ribosomal DNA (rDNA) transcribed by polymerase I is arranged in several hundred tandem repeats, each containing a copy of each of three rRNA genes. The promoter sequences are located in stretches of DNA separating the genes. At the transcriptional start site lies a TATA-like sequence called the *ribosomal initiator element* (rInr). Farther upstream, 150 to 200 bp from the start site, is the *upstream promoter element* (UPE). Both elements aid transcription by binding proteins that recruit RNA polymerase I.

2. *RNA Polymerase II.* Promoters for RNA polymerase II, like prokaryotic promoters, include a set of consensus sequences that define the start site and recruit the polymerase. However, the promoter can contain any combination of a number of possible consensus sequences. Unique to eukaryotes, they also include enhancer elements that can be very distant (more than 1 kb) from the start site.

3. *RNA Polymerase III.* Promoters for RNA polymerase III are *within* the transcribed sequence, downstream of the start site. There are two types of intergenic promoters for RNA polymerase III. Type I promoters, found in the 5S rRNA

gene, contain two short conserved sequences known as the A block and the C block. Type II promoters, found in tRNA genes, consist of two 11-bp sequences, the A block and the B block, situated about 15 bp from either end of the gene.

Three common elements can be found in the RNA polymerase II promoter region

RNA polymerase II transcribes all of the protein-coding genes in eukaryotic cells. Promoters for RNA polymerase II, like those for bacterial polymerases, are generally located on the 5′ side of the start site for transcription. Because these sequences are on the *same* molecule of DNA as the genes being transcribed, they are called *cis-acting elements.* The most commonly recognized cis-acting element for genes transcribed by RNA polymerase II is called the *TATA box* on the basis of its consensus sequence (Figure 29.23). The TATA box is usually found between positions −30 and −100. Note that the eukaryotic TATA box closely resembles the prokaryotic −10 sequence (TATAAT) but is farther from the start site. The mutation of a single base in the TATA box markedly impairs promoter activity. Thus, the precise sequence, not just a high content of AT pairs, is essential.

$5'\ T_{82}\ A_{97}\ T_{93}\ A_{85}\ A_{63}\ A_{88}\ A_{50}\ 3'$

TATA box

FIGURE 29.23 TATA box. Comparisons of the sequences of more than 100 eukaryotic promoters led to the consensus sequence shown. The subscripts denote the frequency (%) of the base at that position.

The TATA box is often paired with an *initiator element* (Inr), a sequence found at the transcriptional start site, between positions −3 and +5. This sequence defines the start site because the other promoter elements are at variable distances from that site. Its presence increases transcriptional activity.

A third element, the *downstream core promoter element* (DPE), is commonly found in conjunction with the Inr in transcripts that lack the TATA box. In contrast with the TATA box, the DPE is found downstream of the start site, between positions +28 and +32.

Additional regulatory sequences are located between −40 and −150. Many promoters contain a *CAAT box,* and some contain a *GC box* (Figure 29.24). Constitutive genes (genes that are continuously expressed rather than regulated) tend to have GC boxes in their promoters. The positions of these upstream sequences vary from one promoter to another, in contrast with the quite constant location of the −35 region in prokaryotes. Another difference is that the CAAT box and the GC box can be effective when present on the template (antisense) strand, unlike the −35 region, which must be present on the coding (sense) strand. These differences between prokaryotes and eukaryotes correspond to fundamentally different mechanisms for the recognition of cis-acting elements. The −10 and −35 sequences in prokaryotic promoters are binding sites for RNA polymerase and its associated σ factor. In contrast, the TATA, CAAT, and GC boxes and other cis-acting elements in eukaryotic promoters are recognized by proteins other than by RNA polymerase itself.

5′ GGNCAATCT 3′

CAAT box

5′ GGGCGG 3′

GC box

FIGURE 29.24 CAAT box and GC box. Consensus sequences for the CAAT and GC boxes of eukaryotic promoters for mRNA precursors.

The TFIID protein complex initiates the assembly of the active transcription complex

Cis-acting elements constitute only part of the puzzle of eukaryotic gene expression. *Transcription factors* that bind to these elements also are required. For example, RNA polymerase II is guided to the start site by a set of transcription factors known collectively as *TFII* (*TF* stands for transcription factor, and *II* refers to RNA polymerase II). Individual TFII factors are called TFIIA, TFIIB, and so on.

In TATA-box promoters, the key initial event is the recognition of the TATA box by the TATA-box-binding protein (TBP), a 30-kDa component of the 700-kDa TFIID complex (Figure 29.25). In TATA-less promoters, other proteins in the TFIID complex bind the core promoter elements but, because less is known about these interactions, we will consider only the

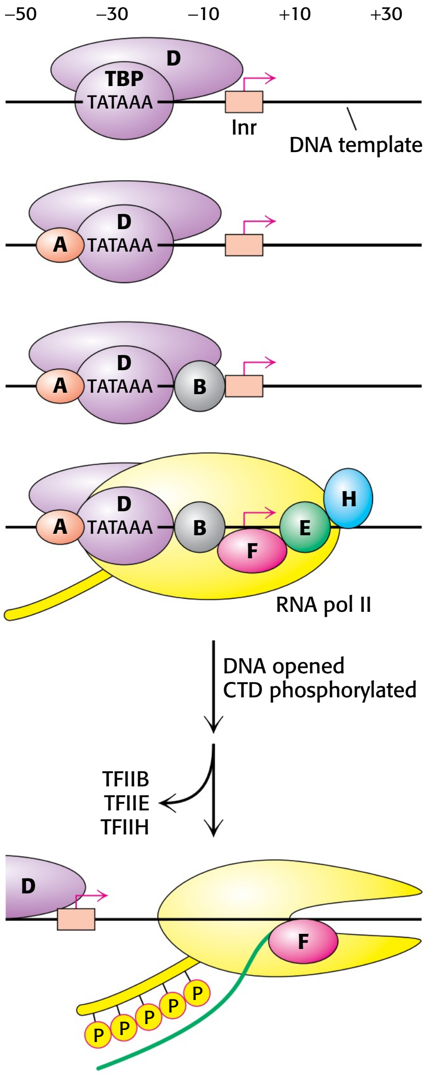


FIGURE 29.25 Transcription initiation. Transcription factors TFIIA, B, D, E, F, and H are essential in initiating transcription by RNA polymerase II. The step-by-step assembly of these general transcription factors begins with the binding of TFIID (purple) to the TATA box. [The TATA-box-binding protein (TBP), a component of TFIID, recognizes the TATA box.] After assembly, TFIIH opens the DNA double helix and phosphorylates the carboxyl-terminal domain (CTD), allowing the polymerase to leave the promoter and begin transcription. The red arrow marks the transcription start site.

TATA-box–TBP binding interaction. TBP binds 10^5 times as tightly to the TATA box as to nonconsensus sequences; the dissociation constant of the TBP–TATA-box complex is approximately 1 nM. TBP is a saddle-shaped protein consisting of two similar domains (Figure 29.26). The TATA box of DNA binds to the concave surface of TBP. This binding induces large conformational changes in the bound DNA. The double helix is substantially unwound to widen its *minor groove,* enabling it to make extensive contact with the antiparallel β strands on the concave side of TBP. Hydrophobic interactions are prominent at this interface. Four phenylalanine residues, for example, are intercalated between base pairs of the TATA box. The flexibility of AT-rich sequences is generally exploited here in bending the DNA. Immediately outside the TATA box, classical B-DNA resumes. The TBP–TATA-box complex is distinctly asymmetric. The asymmetry is crucial for specifying a unique start site and ensuring that transcription proceeds unidirectionally.

TBP bound to the TATA box is the heart of the initiation complex (Figure 29.25). The surface of the TBP saddle provides docking sites for the binding of other components. Additional transcription factors assemble on this nucleus in a defined sequence. TFIIA is recruited, followed by TFIIB; then TFIIF, RNA polymerase II, TFIIE, and TFIIH join the other factors to form a complex called the *basal transcription apparatus*. During the formation of the basal transcription apparatus, the carboxyl-terminal domain (CTD) is unphosphorylated and plays a role in transcription regulation through its binding to an enhancer-associated complex called mediator (Section 32.2). The phosphorylated CTD stabilizes transcription elongation by RNA polymerase II and recruits RNA-processing enzymes that act in the course of elongation. *Phosphorylation of the CTD by TFIIH marks the transition from initiation to elongation.* The importance of the carboxyl-terminal domain is highlighted by the finding that yeast containing mutant polymerase II with fewer than 10 repeats in the CTD are not viable. Most of the factors are released before the polymerase leaves the promoter and can then participate in another round of initiation.

Multiple transcription factors interact with eukaryotic promoters

The basal transcription complex described in the preceding section initiates transcription at a low frequency. Additional transcription factors that bind to other sites are required to achieve a high rate of mRNA synthesis. Their role is to selectively stimulate specific genes. Upstream stimulatory sites in eukaryotic genes are diverse in sequence and variable in position. Their variety suggests that they are recognized by many different specific proteins. Indeed, many transcription factors have been isolated, and their binding sites have been identified by footprinting experiments. For example, *heat-shock transcription factor* (HSTF) is expressed in *Drosophila* after an abrupt increase in temperature. This 93-kDa DNA-binding protein binds to the following consensus sequence:

$$5'\text{—CNNGAANNTCCNNG—}3'$$

Several copies of this sequence, known as the *heat-shock response element,* are present starting at a site 15 bp upstream of the TATA box.

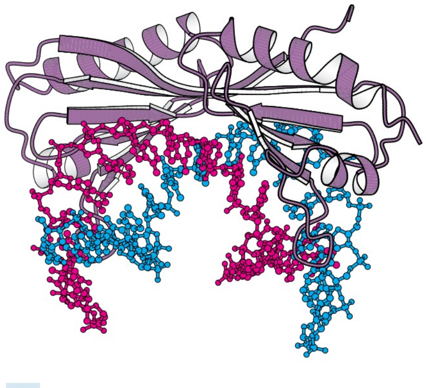
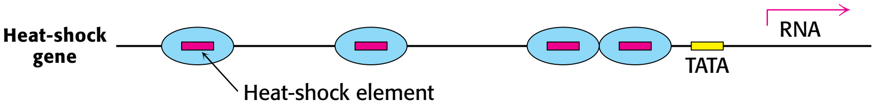


FIGURE 29.26 Complex formed by TATA-box-binding protein and DNA. The saddlelike structure of the protein sits atop a DNA fragment. *Notice* that the DNA is significantly unwound and bent. [Drawn from 1CDW.pdb.]



HSTF differs from σ^{32}, a heat-shock protein of *E. coli* (p. 865), in binding directly to response elements in heat-shock promoters rather than first becoming associated with RNA polymerase.

Enhancer sequences can stimulate transcription at start sites thousands of bases away

The activities of many promoters in higher eukaryotes are greatly increased by another type of cis-acting element called an *enhancer.* Enhancer sequences have no promoter activity of their own *yet can exert their stimulatory actions over distances of several thousand base pairs. They can be upstream, downstream, or even in the midst of a transcribed gene*. Moreover, enhancers are effective when present on *either the coding or non-coding DNA strand*.

A particular enhancer is effective only in certain cells. For example, the immunoglobulin enhancer functions in B lymphocytes but not elsewhere. Cancer can result if the relation between genes and enhancers is disrupted. In Burkitt lymphoma and B-cell leukemia, a chromosomal translocation brings the proto-oncogene *myc* (a transcription factor itself) under the control of a powerful immunoglobin enhancer. The consequent dysregulation of the *myc* gene is believed to play a role in the progression of the cancer.

Transcription factors and other proteins that bind to regulatory sites on DNA can be regarded as passwords that cooperatively open multiple locks, giving RNA polymerase access to specific genes. The discovery of promoters and enhancers has allowed us to gain a better understanding of how genes are selectively expressed in eukaryotic cells. The regulation of gene transcription, discussed in Chapter 32, is the fundamental means of controlling gene expression.

Although bacteria lack TBP, archaea utilize a TBP molecule that is structurally quite similar to the eukaryotic protein. In fact, transcriptional control processes in archaea are, in general, much more similar to those in eukaryotes than are the processes in bacteria. Many components of the eukaryotic transcriptional machinery evolved from those in an ancestor of archaea.

29.3 The Transcription Products of Eukaryotic Polymerases Are Processed

Virtually all the initial products of transcription are further processed in eukaryotes. For example, primary transcripts (pre-mRNA molecules), the products of RNA polymerase II action, acquire a cap at their 5′ ends and a poly(A) tail at their 3′ends. Most importantly, *nearly all mRNA precursors in higher eukaryotes are spliced*. Introns are precisely excised from primary transcripts, and exons are joined to form mature mRNAs with continuous messages. Some mature mRNAs are only a tenth the size of their precursors, which can be as large as 30 kb or more. The pattern of splicing can be regulated in the course of development to generate variations on a theme, such as membrane-bound or secreted forms of antibody molecules. Alternative splicing enlarges the repertoire of proteins in eukaryotes and is one clear illustration of why the proteome is more complex than the genome. The particular processing steps and the factors taking part vary according to the type of RNA polymerase.

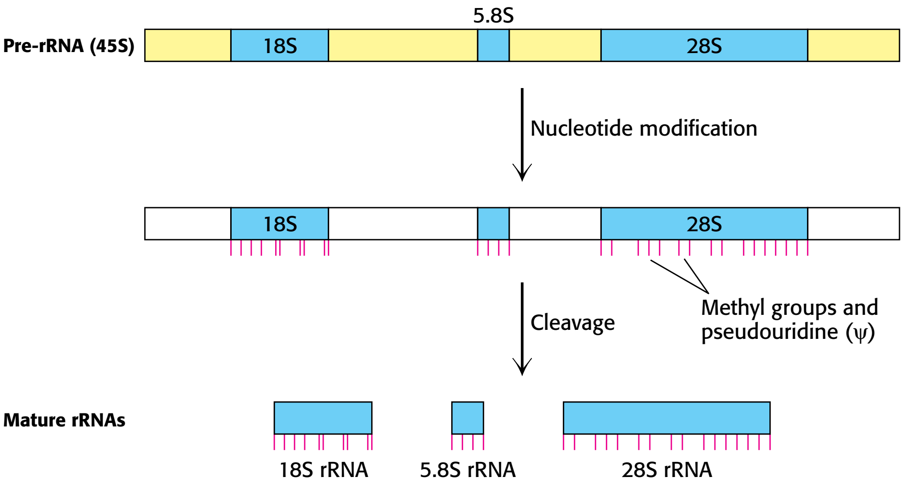


FIGURE 29.27 Processing of eukaryotic pre-rRNA. The mammalian pre-rRNA transcript contains the RNA sequences destined to become the 18S, 5.8S, and 28S rRNAs of the small and large ribosomal subunits. First, nucleotides are modified: small nucleolar ribonucleoproteins methylate specific nucleoside groups and convert selected uridines into pseudouridines (indicated by red lines). Next, the pre-rRNA is cleaved and packaged to form mature ribosomes, in a highly regulated process in which more than 200 proteins take part.

RNA polymerase I produces three ribosomal RNAs

Several RNA molecules are key components of ribosomes. RNA polymerase I transcription produces a single precursor (45S in mammals) that encodes three RNA components of the ribosome: the 18S rRNA, the 28S rRNA, and the 5.8S rRNA (Figure 29.27). The 18S rRNA is the RNA component of the small ribosomal subunit (40S), and the 28S and 5.8S rRNAs are two RNA components of the large ribosomal subunit (60S). The other RNA component of the large ribosomal subunit, the 5S rRNA, is transcribed by RNA polymerase III as a separate transcript.

The cleavage of the precursor into three separate rRNAs is actually the final step in its processing. First, the nucleotides of the pre-rRNA sequences destined for the ribosome undergo extensive modification, on both ribose and base components, directed by many *small nucleolar ribonucleoproteins* (snoRNPs), each of which consists of one snoRNA and several proteins. The pre-rRNA is assembled with ribosomal proteins, as guided by processing factors, in a large ribonucleoprotein. For instance, the small-subunit (SSU) processome is required for 18S rRNA synthesis and can be visualized in electron micrographs as a terminal knob at the 5′ ends of the nascent rRNAs (Fig. 29.28). Finally, rRNA cleavage (sometimes coupled with additional processing steps) releases the mature rRNAs assembled with ribosomal proteins as ribosomes. Like those of RNA polymerase I transcription itself, most of these processing steps take place in the cell nucleolus, a nuclear subcompartment.

RNA polymerase III produces transfer RNA

Eukaryotic tRNA transcripts are among the most processed of all RNA polymerase III transcripts. Like those of prokaryotic tRNAs, the 5′ leader is cleaved by RNase P, the 3′ trailer is removed, and CCA is added by the CCA-adding enzyme (Figure 29.29). Eukaryotic tRNAs are also heavily modified on base and ribose moieties; these modifications are important for function. In contrast with prokaryotic tRNAs, many eukaryotic pre-tRNAs are also spliced by an endonuclease and a ligase to remove an intron.

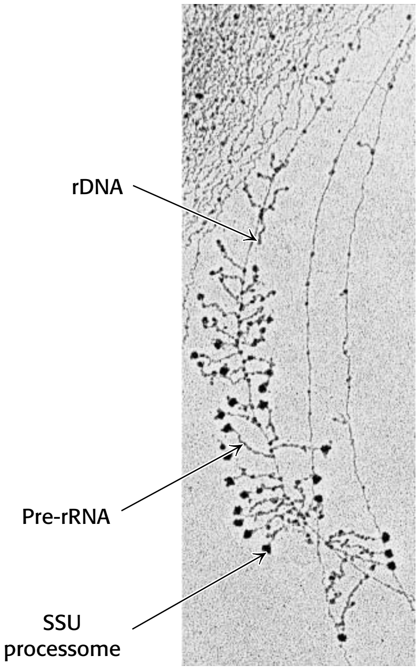


FIGURE 29.28 Visualization of rRNA transcription and processing in eukaryotes. Transcription of rRNA and its assembly into precursor ribosomes can be visualized by electron microscopy. The structures resemble Christmas trees: the trunk is the rDNA and each branch is a pre-rRNA transcript. Transcription starts at the top of the tree, where the shortest transcripts can be seen, and progresses down the rDNA to the end of the gene. The terminal knobs visible at the end of some pre-rRNA transcripts likely correspond to the SSU processome, a large ribonucleoprotein required for processing the pre-rRNA. [Reprinted by permission from Macmillan Publishers Ltd: *Nature* 417:967-970, F. Dragon et. al. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis, © 2002.]

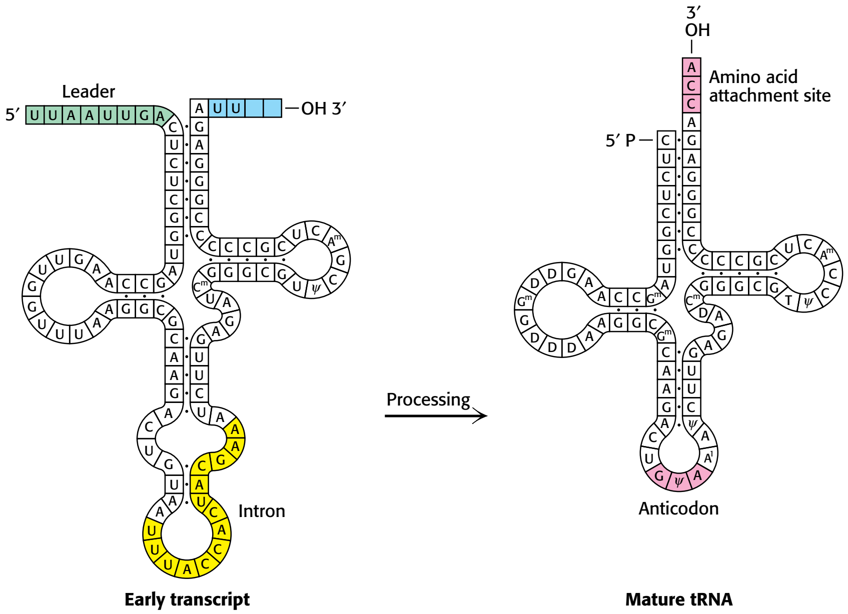

Leader
5′ UUAAUUGA
AUU
—OH 3′
Intron
Early transcript
Processing
3′
OH
Amino acid
attachment site
5′ P
Anticodon
Mature tRNA

FIGURE 29.29 Transfer RNA precursor processing. The conversion of a yeast tRNA precursor into a mature tRNA requires the removal of a 14-nucleotide intron (yellow), the cleavage of a 5′ leader (green), and the removal of UU and the attachment of CCA at the 3′ end (red). In addition, several bases are modified.

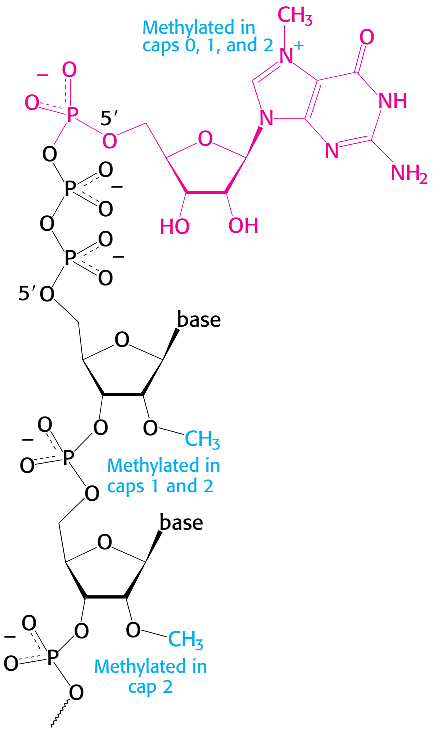
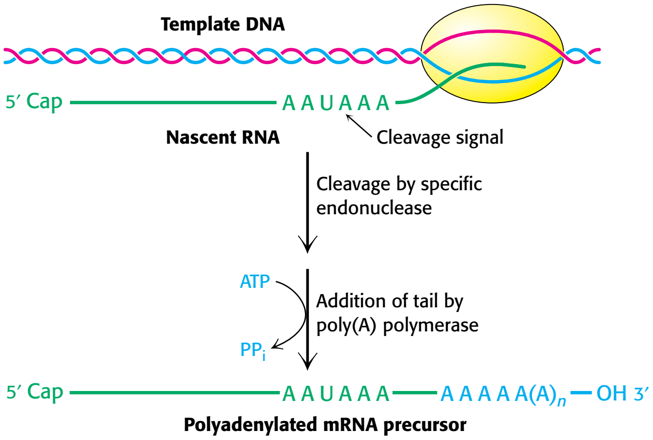

Methylated in
caps 0, 1, and 2
CH3
5′
5′
base
CH3
Methylated in
caps 1 and 2
base
CH3
Methylated in
cap 2
HO
OH
NH
NH2

FIGURE 29.30 Capping the 5′ end. Caps at the 5′ end of eukaryotic mRNA include 7-methylguanylate (red) attached by a triphosphate linkage to the ribose at the 5′ end. None of the riboses are methylated in cap 0, one is methylated in cap 1, and both are methylated in cap 2.

FIGURE 29.31 Polyadenylation of a primary transcript. A specific endonuclease cleaves the RNA downstream of AAUAAA. Poly(A) polymerase then adds about 250 adenylate residues.


Template DNA
5′ Cap
AAUAAA
Nascent RNA
Cleavage signal
Cleavage by specific endonuclease
ATP
PPi
Addition of tail by poly(A) polymerase
5′ Cap
AAUAAA
AAAAA(A)n
OH 3′
Polyadenylated mRNA precursor

The product of RNA polymerase II, the pre-mRNA transcript, acquires a 5′ cap and a 3′ poly(A) tail

Perhaps the most extensively studied transcription product is the product of RNA polymerase II: most of this RNA will be processed to mRNA. The immediate product of RNA polymerase II is sometimes referred to as precursor-to-messenger RNA, or *pre-mRNA*. Most pre-mRNA molecules are spliced to remove the introns. Moreover, both the 5′ and the 3′ ends are modified, and both modifications are retained as the pre-mRNA is converted into mRNA.

As in prokaryotes, eukaryotic transcription usually begins with A or G. However, the 5′ triphosphate end of the nascent RNA chain is immediately modified. First, a phosphoryl group is released by hydrolysis. The diphosphate 5′ end then attacks the α-phosphorus atom of GTP to form a very unusual 5′–5′ triphosphate linkage. This distinctive terminus is called a *cap* (Figure 29.30). The N-7 nitrogen of the terminal guanine is then methylated by *S*-adenosylmethionine to form *cap 0*. The adjacent riboses may be methylated to form *cap 1* or *cap 2*. Transfer RNA and ribosomal RNA molecules, in contrast with messenger RNAs and with small RNAs that participate in splicing, do not have caps. Caps contribute to the stability of mRNAs by protecting their 5′ ends from phosphatases and nucleases. In addition, caps enhance the translation of mRNA by eukaryotic protein-synthesizing systems.

As mentioned earlier, pre-mRNA is also modified at the 3′ end. *Most eukaryotic mRNAs contain a polyadenylate, poly(A), tail at that end,* added after transcription has ended. The DNA template does not encode this poly(A) tail. Indeed, the nucleotide preceding poly(A) is not the last nucleotide to be transcribed. Some primary transcripts contain hundreds of nucleotides beyond the 3′ end of the mature mRNA.

How is the 3′ end of the pre-mRNA given its final form? *Eukaryotic primary transcripts are cleaved by a specific endonuclease that recognizes the sequence AAUAAA* (Figure 29.31).

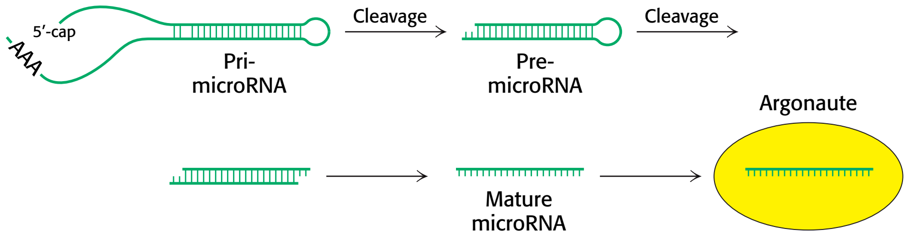


FIGURE 29.32 Small regulatory RNA production. A pathway from a transcription product including a microRNA to the mature microRNA bound to an Argonaute protein. The initial transcription product, a pri-microRNA, is first cleaved to a small double-stranded RNA called a pre-microRNA. One of the strands of the pre-microRNA, the mature microRNA, is then bound by an Argonaute protein.

Cleavage does not take place if this sequence or a segment of some 20 nucleotides on its 3′ side is deleted. The presence of internal AAUAAA sequences in some mature mRNAs indicates that AAUAAA is only part of the cleavage signal; its context also is important. After cleavage of the pre-RNA by the endonuclease, a *poly(A) polymerase* adds about 250 adenylate residues to the 3′ end of the transcript; ATP is the donor in this reaction.

The role of the poly(A) tail is still not firmly established despite much effort. However, evidence is accumulating that it enhances translation efficiency and the stability of mRNA. Blocking the synthesis of the poly(A) tail by exposure to *3′-deoxyadenosine (cordycepin)* does not interfere with the synthesis of the primary transcript. Messenger RNA devoid of a poly(A) tail can be transported out of the nucleus. However, an mRNA molecule devoid of a poly(A) tail is usually a much less effective template for protein synthesis than is one with a poly(A) tail. Indeed, some mRNAs are stored in an unadenylated form and receive the poly(A) tail only when translation is imminent. The half-life of an mRNA molecule may be determined in part by the rate of degradation of its poly(A) tail.

Small regulatory RNAs are cleaved from larger precursors

Cleavage plays a role in the processing of small single-stranded RNAs (approximately 20–23 nucleotides) called *microRNAs*. MicroRNAs play key roles in gene regulation in eukaryotes, as we shall see in Chapter 32. They are generated from initial transcripts produced by RNA polymerase II and, in some cases, RNA polymerase III. These transcripts fold into hairpin structures that are cleaved by specific nucleases at various stages (Figure 29.32). The final single-stranded RNAs are bound by members of the Argonaute family of proteins where the RNAs help target the regulation of specific genes.

RNA editing changes the proteins encoded by mRNA

Remarkably, the amino acid sequence information encoded by some mRNAs is altered after transcription. *RNA editing* is the term for a change in the nucleotide sequence of RNA after transcription by processes other than RNA splicing. RNA editing is prominent in some systems already discussed. *Apolipoprotein B* (apo B) plays an important role in the transport of triacylglycerols and cholesterol by forming an amphipathic spherical shell around the lipids carried in lipoprotein particles (Section 26.3). Apo B exists in two forms, a 512-kDa *apo B-100* and a 240-kDa *apo B-48*. The larger form, synthesized by the liver, participates in the transport of lipids synthesized in the cell. The smaller form, synthesized by the small intestine, carries dietary fat in the form of chylomicrons. Apo B-48 contains the 2152 N-terminal residues of the 4536-residue apo B-100. This truncated molecule can form lipoprotein particles but cannot bind to the low-density-lipoprotein receptor on cell surfaces. What is the relationship between these two forms of apo B? Experiments revealed that a totally unexpected mechanism for generating diversity is at work: *the changing of the nucleotide sequence of mRNA after its synthesis* (Figure 29.33). *A specific cytidine residue of mRNA is deaminated to uridine, which changes the codon at residue 2153*

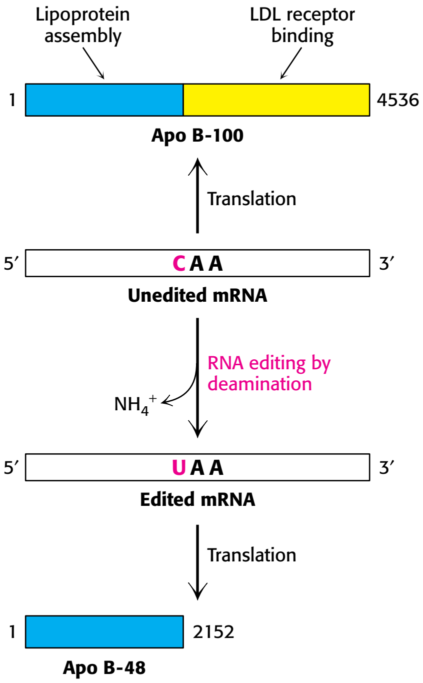


FIGURE 29.33 RNA editing. Enzyme-catalyzed deamination of a specific cytidine residue in the mRNA for apolipoprotein B-100 changes a codon for glutamine (CAA) to a stop codon (UAA). Apolipoprotein B-48, a truncated version of the protein lacking the LDL receptor-binding domain, is generated by this posttranscriptional change in the mRNA sequence. [Information from P. Hodges and J. Scott, *Trends Biochem. Sci.* 17:77, 1992.]

from CAA (Gln) to UAA (stop). The deaminase that catalyzes this reaction is present in the small intestine, but not in the liver, and is expressed only at certain developmental stages.

RNA editing is not confined to apolipoprotein B. Glutamate opens cation-specific channels in the vertebrate central nervous system by binding to receptors in postsynaptic membranes. RNA editing changes a single glutamine codon (CAG) in the mRNA for the glutamate receptor to the codon for arginine (read as CGG). The substitution of Arg for Gln in the receptor prevents Ca^{2+}, but not Na^{+}, from flowing through the channel.

RNA editing is likely much more common than was formerly thought. The chemical reactivity of nucleotide bases, including the susceptibility to deamination that necessitates complex DNA-repair mechanisms, has been harnessed as an engine for generating molecular diversity at the RNA and, hence, protein levels.

In trypanosomes (parasitic protozoans), a different kind of RNA editing markedly changes several mitochondrial mRNAs. Nearly half the uridine residues in these mRNAs are inserted by RNA editing. A *guide RNA molecule* identifies the sequences to be modified, and a *poly(U) tail* on the guide donates uridine residues to the mRNAs undergoing editing. DNA sequences evidently do not always faithfully disclose the sequence of encoded proteins: functionally crucial changes to mRNA can take place.

Sequences at the ends of introns specify splice sites in mRNA precursors

Most genes in higher eukaryotes are composed of exons and introns (Section 4.7). The introns must be excised and the exons must be linked to form the final mRNA in a process called *RNA splicing*. This splicing must be exquisitely sensitive: splicing just one nucleotide upstream or downstream of the intended site would create a one-nucleotide shift, which would alter the reading frame on the 3′ side of the splice to give an entirely different amino acid sequence, likely including a premature stop codon. Thus, the correct splice site must be clearly marked. Does a particular sequence denote the splice site? The sequences of thousands of intron–exon junctions within RNA transcripts are known. In eukaryotes from yeast to mammals, these sequences have a common structural motif: *the intron begins with GU and ends with AG*. The consensus sequence at the 5′ splice in vertebrates is AGGUAAGU, where the GU is invariant (Figure 29.34). At the 3′ end of an intron, the consensus sequence is a stretch of *10 pyrimidines* (U or C; termed the *polypyrimidine tract*), followed by any base and then by C, and ending with the invariant AG. Introns also have an important internal site located between 20 and 50 nucleotides upstream of the 3′ splice site; it is called the *branch site* for reasons that will be evident shortly. In yeast, the branch-site sequence is nearly always UACUAAC, whereas, in mammals, a variety of sequences are found.

The 5′ and 3′ splice sites and the branch site are essential for determining where splicing takes place. Mutations in each of these three critical regions lead to aberrant splicing. Introns vary in length from 50 to 10,000 nucleotides, and so the splicing machinery may have to find the 3′ site several

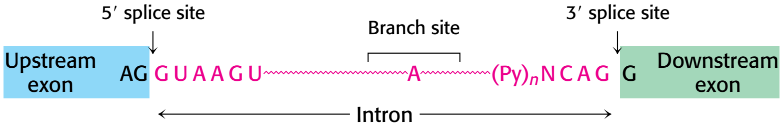


FIGURE 29.34 Splice sites. Consensus sequences for the 5′ splice site and the 3′ splice site are shown. Py stands for pyrimidine.

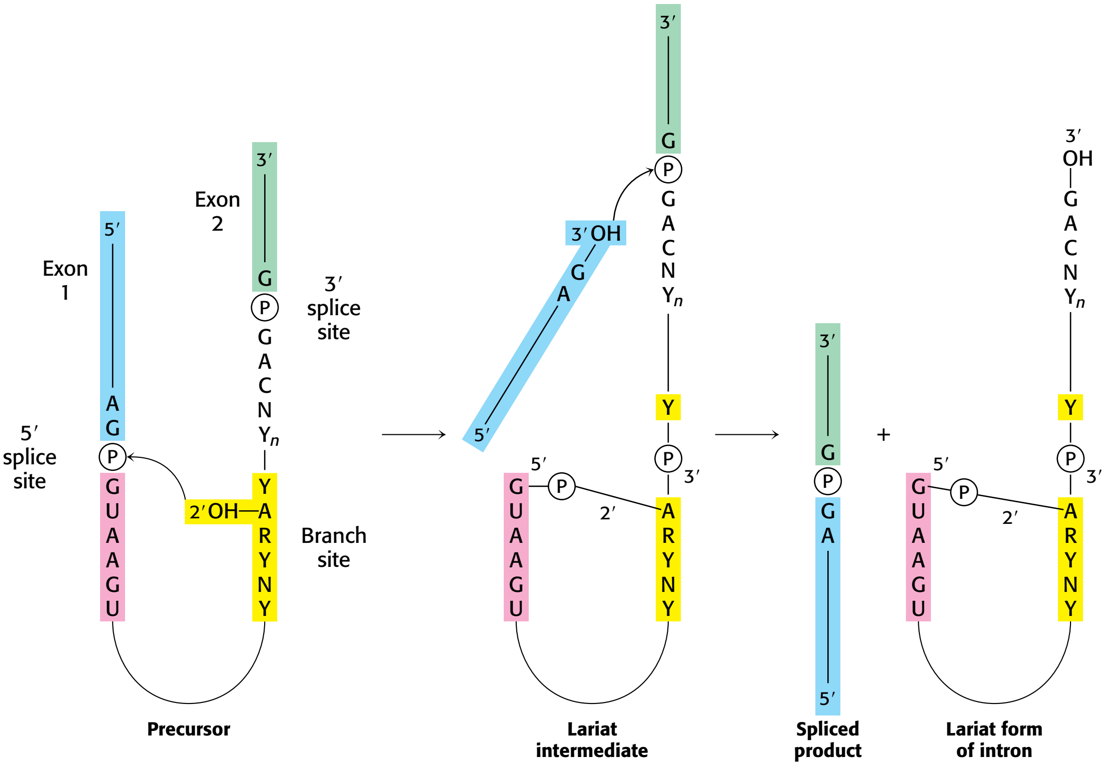
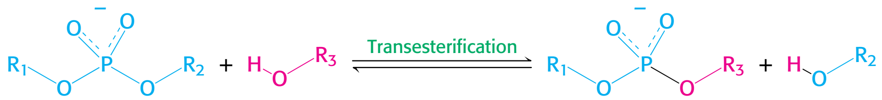


FIGURE 29.35 **Splicing mechanism used for mRNA precursors.** The upstream (5′) exon is shown in blue, the downstream (3′) exon in green, and the branch site in yellow. Y stands for a pyrimidine nucleotide, R for a purine nucleotide, and N for any nucleotide. The 5′ splice site is attacked by the 2′-OH group of the branch-site adenosine residue. The 3′ splice site is attacked by the newly formed 3′-OH group of the upstream exon. The exons are joined, and the intron is released in the form of a lariat. [Information from P. A. Sharp, *Cell* 42:397–408, 1985.]

thousand nucleotides away. Specific sequences near the splice sites (in both the introns and the exons) play an important role in splicing regulation, particularly in designating splice sites when there are many alternatives. Researchers are currently attempting to determine the factors that contribute to splice-site selection for individual mRNAs. Despite our knowledge of splice-site sequences, predicting pre-mRNAs and their protein products from genomic DNA sequence information remains a challenge.

Splicing consists of two sequential transesterification reactions

The splicing of nascent mRNA molecules is a complicated process. It requires the cooperation of several small RNAs and proteins that form a large complex called a *spliceosome*. However, the chemistry of the splicing process is simple. Splicing begins with the cleavage of the phosphodiester bond between the upstream exon (exon 1) and the 5′ end of the intron (Figure 29.35). The attacking group in this reaction is the 2′-OH group of an adenylate residue in the branch site. A 2′–5′ phosphodiester bond is formed between this A residue and the 5′ terminal phosphate of the intron. This reaction is a transesterification.



Note that this adenylate residue is also joined to two other nucleotides by normal 3′—5′ phosphodiester bonds (Figure 29.36). Hence a *branch* is generated at this site, and a *lariat intermediate* is formed.

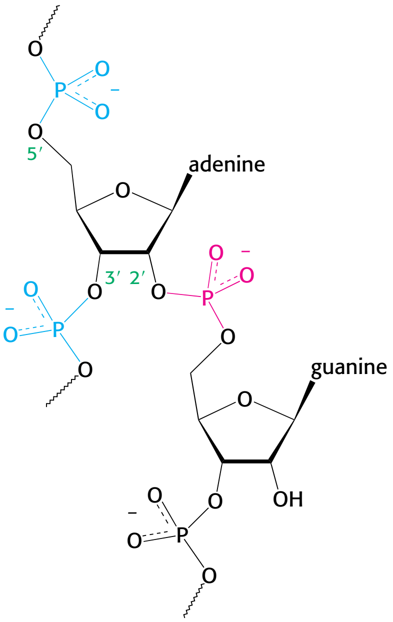


FIGURE 29.36 **Splicing branch point.** The structure of the branch point in the lariat intermediate in which the adenylate residue is joined to three nucleotides by phosphodiester bonds. The new 2′-to-5′ linkage is shown in red, and the usual 3′-to-5′ linkages are shown in blue.

TABLE 29.3 Small nuclear ribonucleoprotein particles (snRNPs) in the splicing of mRNA precursors

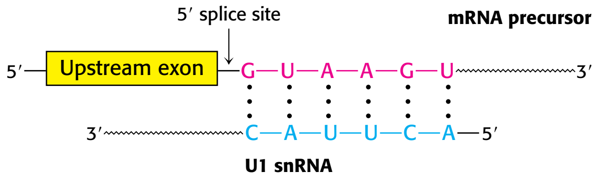
snRNP	Size of snRNA (nucleotides)	Role
U1	165	Binds the 5′ splice site
U2	185	Binds the branch site and forms part of the catalytic center
U5	116	Binds the 5′ splice site and then the 3′ splice site
U4	145	Masks the catalytic activity of U6
U6	106	Catalyzes splicing

The 3′-OH terminus of exon 1 then attacks the phosphodiester bond between the intron and exon 2. Exons 1 and 2 become joined, and the intron is released in lariat form. Again, this reaction is a transesterification. Splicing is thus accomplished by two *transesterification reactions* rather than by hydrolysis followed by ligation. The first reaction generates a free 3′-OH group at the 3′ end of exon 1, and the second reaction links this group to the 5′-phosphate of exon 2. *The number of phosphodiester bonds stays the same during these steps,* which is crucial because it allows the splicing reaction itself to proceed without an energy source such as ATP or GTP.

Small nuclear RNAs in spliceosomes catalyze the splicing of mRNA precursors

The nucleus contains many types of small RNA molecules with fewer than 300 nucleotides, referred to as *snRNAs* (small nuclear RNAs). A few of them—designated U1, U2, U4, U5, and U6—are essential for splicing mRNA precursors. The secondary structures of these RNAs are highly conserved in organisms ranging from yeast to human beings. These RNA molecules are associated with specific proteins to form complexes termed *snRNPs* (*s*mall *n*uclear *r*ibo*n*ucleo*p*rotein particles); investigators often speak of them as "snurps." Spliceosomes are large (60S) dynamic assemblies composed of snRNPs, hundreds of other proteins called *splicing factors,* and the mRNA precursors being processed (Table 29.3).

In mammalian cells, splicing begins with the recognition of the 5′ splice site by the U1 snRNP (Figure 29.37). U1 snRNA contains a highly conserved six-nucleotide sequence, not covered by protein in the snRNP, that base-pairs to the 5′ splice site of the pre-mRNA. This binding initiates spliceosome assembly on the pre-mRNA molecule.



U2 snRNP then binds the branch site in the intron by base-pairing between a highly conserved sequence in U2 snRNA and the pre-mRNA.

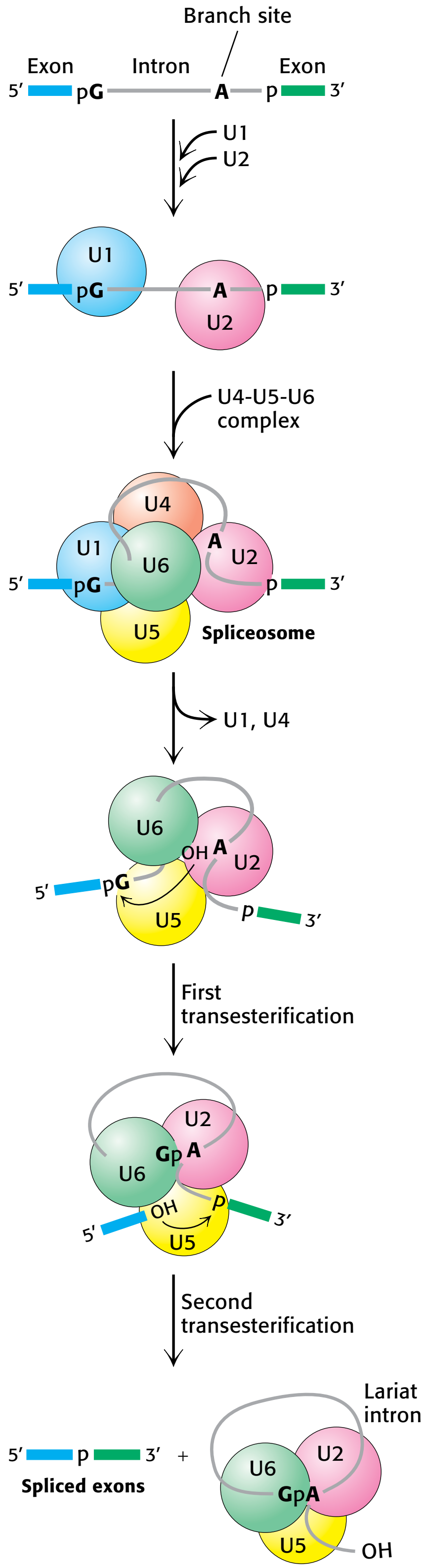


FIGURE 29.37 Spliceosome assembly and action. U1 binds the 5′ splice site and U2 binds to the branch point. A preformed U4-U5-U6 complex then joins the assembly to form the complete spliceosome. The U6 snRNA re-folds and binds the 5′ splice site, displacing U1. Extensive interactions between U6 and U2 displace U4. Then, in the first transesterification step, the branch-site adenosine attacks the 5′ splice site, making a lariat intermediate. U5 holds the two exons in close proximity, and the second transesterification takes place, with the 5′ splice-site hydroxyl group attacking the 3′ splice site. These reactions result in the mature spliced mRNA and a lariat form of the intron bound by U2, U5, and U6. [Information from T. Villa, J. A. Pleiss, and C. Guthrie, *Cell* 109:149–152, 2002.]

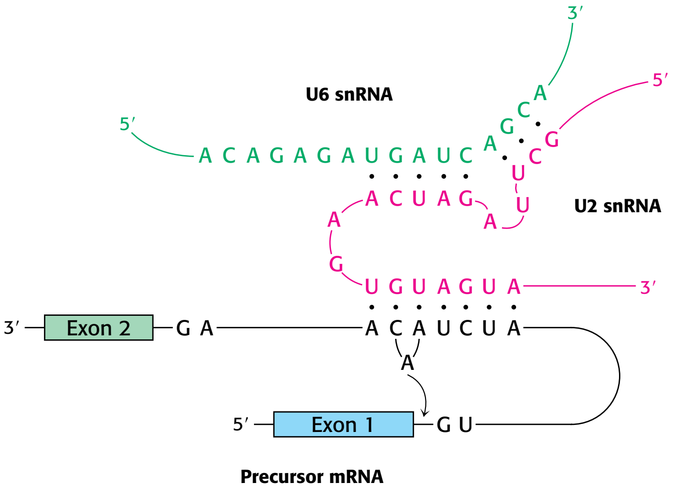


FIGURE 29.38 Splicing catalytic center. The catalytic center of the spliceosome is formed by U2 snRNA (red) and U6 snRNA (green), which are base paired. U2 is also base paired to the branch site of the mRNA precursor. [Information from H. D. Madhani and C. Guthrie, *Cell* 71:803–817, 1992.]

U2 snRNP binding requires ATP hydrolysis. A preassembled U4-U5-U6 tri-snRNP joins this complex of U1, U2, and the mRNA precursor to form the spliceosome. This association also requires ATP hydrolysis.

A revealing view of the interplay of RNA molecules in this assembly came from examining the pattern of cross-links formed by *psoralen,* a reagent that joins neighboring pyrimidines in base-paired regions on treatment with light. These cross-links suggest that splicing takes place in the following way. First, U5 interacts with exon sequences in the 5′ splice site and subsequently with the 3′ exon. Next, U6 disengages from U4 and undergoes an intramolecular rearrangement that permits base-pairing with U2 as well as interaction with the 5′ end of the intron, displacing U1 from the spliceosome. The U2-U6 helix is indispensable for splicing, suggesting that *U2 and U6 snRNAs probably form the catalytic center of the spliceosome* (Figure 29.38). U4 serves as an inhibitor that masks U6 until the specific splice sites are aligned. These rearrangements result in the first transesterification reaction, cleaving the 5′ exon and generating the lariat intermediate.

Further rearrangements of RNA in the spliceosome facilitate the second transesterification. In these rearrangements, U5 aligns the free 5′ exon with the 3′ exon such that the 3′-hydroxyl group of the 5′ exon is positioned to nucleophilically attack the 3′ splice site to generate the spliced product. U2, U5, and U6 bound to the excised lariat intron are released to complete the splicing reaction.

Many of the steps in the splicing process require ATP hydrolysis. How is the free energy associated with ATP hydrolysis used to power splicing? To achieve the well-ordered rearrangements necessary for splicing, ATP-powered RNA helicases must unwind RNA helices and allow alternative base-pairing arrangements to form. Thus, two features of the splicing process are noteworthy. First, *RNA molecules play key roles in directing the alignment of splice sites and in carrying out catalysis*. Second, *ATP-powered helicases unwind RNA duplex intermediates that facilitate catalysis and induce the release of snRNPs from the mRNA.*

Transcription and processing of mRNA are coupled

Although the transcription and processing of mRNAs have been described herein as separate events in gene expression, experimental evidence suggests that the two steps are coordinated by the carboxyl-terminal domain of RNA polymerase II. We have seen that the CTD consists of a unique repeated seven-amino-acid sequence, YSPTSPS. Either S_2 or S_5 or both may be phosphorylated in the various repeats. The phosphorylation state of the CTD is controlled by a number of kinases and phosphatases and leads the CTD to bind many of

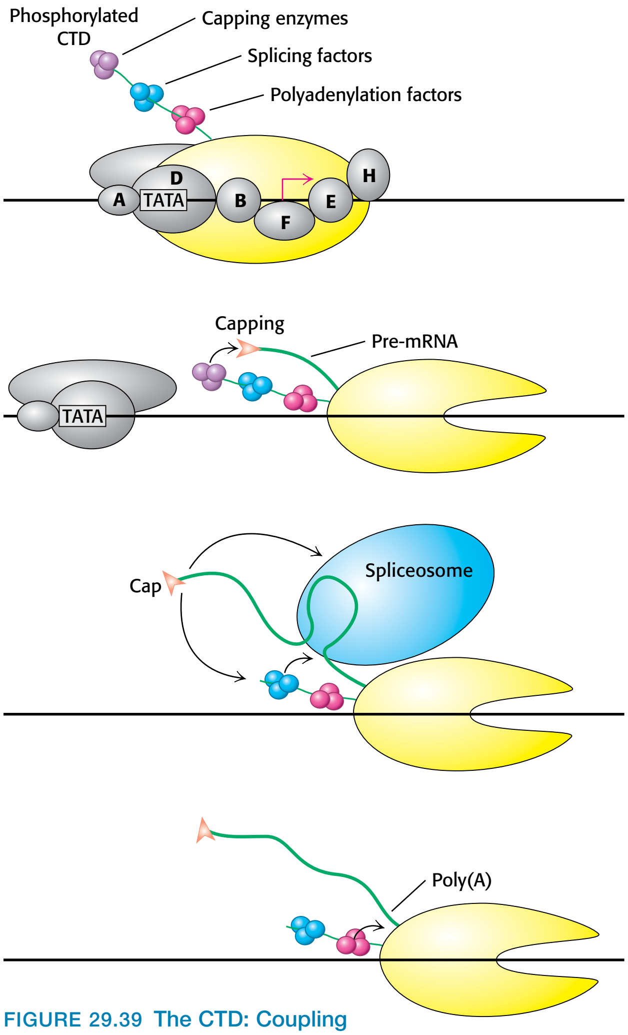


FIGURE 29.39 The CTD: Coupling transcription to pre-mRNA processing. The transcription factor TFIIH phosphorylates the carboxyl-terminal domain (CTD) of RNA polymerase II, signaling the transition from transcription initiation to elongation. The phosphorylated CTD binds factors required for pre-mRNA capping, splicing, and polyadenylation. These proteins are brought in close proximity to their sites of action on the nascent pre-mRNA as it is transcribed during elongation. [Information from P. A. Sharp, *TIBS* 30:279–281, 2005.]

the proteins having roles in RNA transcription and processing. The CTD contributes to efficient transcription by recruiting these proteins to the pre-mRNA (Figure 29.39), including:

1. capping enzymes, which methylate the 5′ guanine on the pre-mRNA immediately after transcription begins;
2. components of the splicing machinery, which initiate the excision of each intron as it is synthesized; and
3. an endonuclease that cleaves the transcript at the poly(A) addition site, creating a free 3′-OH group that is the target for 3′ adenylation.

These events take place sequentially, directed by the phosphorylation state of the CTD.

Mutations that affect pre-mRNA splicing cause disease

Mutations in either the pre-mRNA (cis-acting) or the splicing factors (trans-acting) can cause defective pre-mRNA splicing. Mutations in the pre-mRNA cause some forms of thalassemia, a group of hereditary anemias characterized by the defective synthesis of hemoglobin (Section 7.4). Cis-acting mutations that cause aberrant splicing can occur at the 5′ or 3′ splice sites in either of the two introns of the hemoglobin β chain or in its exons. The mutations usually result in an incorrectly spliced pre-mRNA that, because of a premature stop codon, cannot encode a full-length protein. The defective mRNA is normally degraded rather than translated. Mutations in the 5′ splice site may alter that site such that the splicing machinery cannot recognize it, forcing the machinery to find another 5′ splice site in the intron and introducing the potential for a premature stop codon. Mutations in the intron itself may create a new 5′ splice site; in this case, either one of the two splice sites may be recognized (Figure 29.40). Consequently, some normal protein can be made, and so the disease is less severe. *Mutations affecting splicing have been estimated to cause at least 15% of all genetic diseases.*

Disease-causing mutations may also appear in splicing factors. Retinitis pigmentosa is a disease of acquired blindness, first described in 1857, with an incidence of 1/3500. About 5% of the autosomal dominant form of retinitis

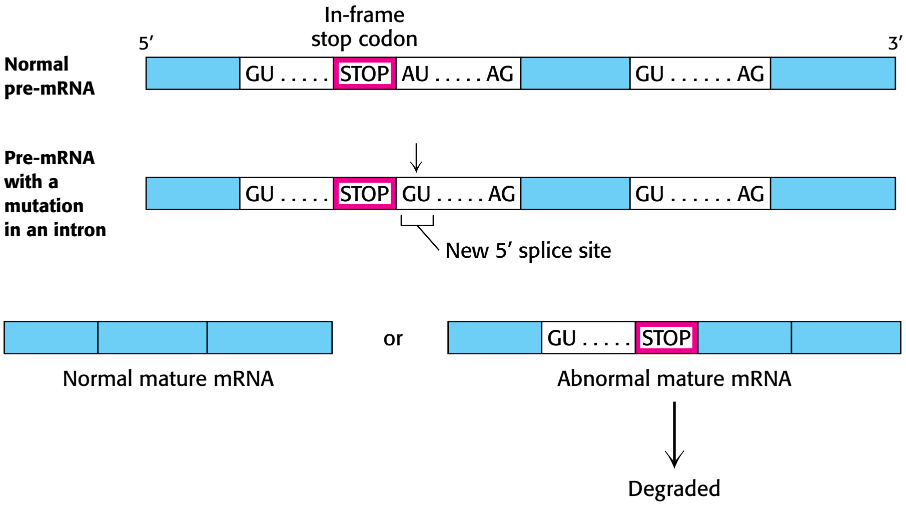


FIGURE 29.40 A splicing mutation that causes thalassemia. An A-to-G mutation within the first intron of the gene for the human hemoglobin β chain creates a new 5′ splice site (GU). Both 5′ splice sites are recognized by the U1 snRNP; thus splicing may sometimes create a normal mature mRNA and an abnormal mature mRNA that contains intron sequences. The normal mature mRNA is translated into a hemoglobin β chain. Because it includes intron sequences, the abnormal mature mRNA now has a premature stop codon and is degraded.

pigmentosa is likely due to mutations in the hPrp8 protein, a pre-mRNA splicing factor that is a component of the U4-U5-U6 tri-snRNP. How a mutation in a splicing factor that is present in all cells causes disease only in the retina is not clear; nevertheless, retinitis pigmentosa is a good example of how mutations that disrupt spliceosome function can cause disease.

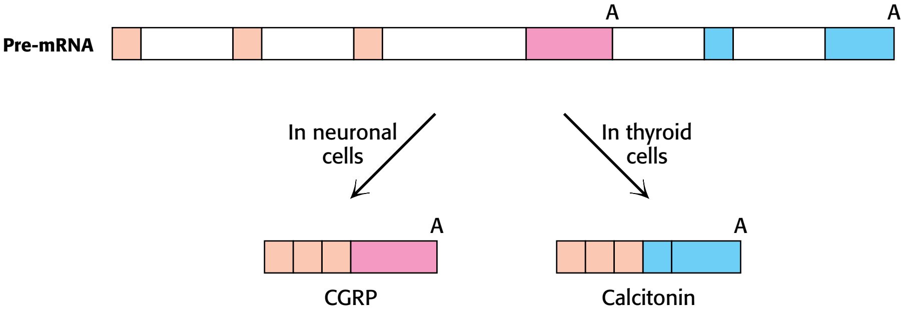


FIGURE 29.41 An example of alternative splicing. In human beings, two very different hormones are produced from a single calcitonin/CGRP pre-mRNA. Alternative splicing produces the mature mRNA for either calcitonin or CGRP (calcitonin-gene-related protein), depending on the cell type in which the gene is expressed. Each alternative transcript incorporates one of two alternative polyadenylation signals (A) present in the pre-mRNA.

Most human pre-mRNAs can be spliced in alternative ways to yield different proteins

Alternative splicing is a widespread mechanism for generating protein diversity. Different combinations of exons from the same gene may be spliced into a mature RNA, producing distinct forms of a protein for specific tissues, developmental stages, or signaling pathways. What controls which splicing sites are selected? The selection is determined by the binding of trans-acting splicing factors to cis-acting sequences in the pre-mRNA. Most alternative splicing leads to changes in the coding sequence, resulting in proteins with different functions. *Alternative splicing provides a powerful mechanism for expanding the versatility of genomic sequences through combinatorial control.* Consider a gene with five positions at which splicing can take place. With the assumption that these alternative splicing pathways can be regulated independently, a total of $2^5 = 32$ different mRNAs can be generated.

Sequencing of the human genome has revealed that most pre-mRNAs are alternatively spliced, leading to a much greater number of proteins than would be predicted from the number of genes. An example of alternative splicing leading to the expression of two different proteins, each in a different tissue, is provided by the gene encoding both calcitonin and calcitonin-gene-related peptide (CGRP; Figure 29.41). In the thyroid gland, the inclusion of exon 4 in one splicing pathway produces calcitonin, a peptide hormone that regulates calcium and phosphorus metabolism. In neuronal cells, the exclusion of exon 4 in another splicing pathway produces CGRP, a peptide hormone that acts as a vasodilator. A single pre-mRNA thus yields two very different peptide hormones, depending on cell type. In this case, only two proteins result from alternative splicing; however, in other cases, many more can be produced. An extreme example is the *Drosophila* pre-mRNA that encodes DSCAM, a neuronal protein affecting axon connectivity. Alternative splicing of this pre-mRNA has the potential to produce 38,016 different combinations of exons, a greater number than the total number of genes in the *Drosophila* genome. However, only a fraction of these potential mRNAs appear to be produced, owing to regulatory mechanisms that are not yet well understood. Several human diseases that can be attributed to defects in alternative splicing are listed in Table 29.4.

TABLE 29.4 Selected human diseases attributed to defects in alternative splicing

Disorder	Gene or its product
Acute intermittent porphyria	Porphobilinogen deaminase
Breast and ovarian cancer	*BRCA*1
Cystic fibrosis	CFTR
Frontotemporal dementia	τ protein
Hemophilia A	Factor VIII
HGPRT deficiency (Lesch–Nyhan syndrome)	Hypoxanthine-guanine phosphoribosyltransferase
Leigh encephalomyelopathy	Pyruvate dehydrogenase E1α
Severe combined immunodeficiency	Adenosine deaminase
Spinal muscle atrophy	*SMN*1 or *SMN*2

Further understanding of alternative splicing and the mechanisms of splice-site selection will be crucial to understanding how the proteome represented by the human genome is expressed.

29.4 The Discovery of Catalytic RNA Was Revealing in Regard to Both Mechanism and Evolution

RNAs form a surprisingly versatile class of molecules. As we have seen, splicing is catalyzed largely by RNA molecules, with proteins playing a secondary role. Another enzyme that contains a key RNA component is ribonuclease P, which catalyzes the maturation of tRNA by endonucleolytic cleavage of nucleotides from the 5′ end of the precursor molecule. Finally, as we shall see in Chapter 30, the RNA component of ribosomes is the catalyst that carries out protein synthesis.

The versatility of RNA first became clear from observations of the processing of ribosomal RNA in a single-cell eukaryote. In *Tetrahymena* (a ciliated protozoan), a 414-nucleotide intron is removed from a 6.4-kb precursor to yield the mature 26S rRNA molecule (Figure 29.42). In an elegant series of studies of this splicing reaction, Thomas Cech and his coworkers established that the RNA spliced itself to precisely excise the intron. These remarkable experiments demonstrated that an RNA molecule can *splice itself* in the absence of protein. Indeed, the RNA alone is catalytic and, under certain conditions, is thus a *ribozyme*. More than 1500 similar introns have since been found in species as widely dispersed as bacteria and eukaryotes, though not in vertebrates. Collectively, they are referred to as *group I introns*.

The *self-splicing* reaction in the group I intron requires an added guanosine nucleotide. Nucleotides were originally included in the reaction mixture because it was thought that ATP or GTP might be needed as an energy source. Instead, the nucleotides were found to be necessary as cofactors. The required cofactor proved to be a guanosine unit, in the form of guanosine, GMP, GDP, or GTP. G (denoting any one of these species) serves not as an energy source but as an attacking group that becomes transiently incorporated into the RNA (Figure 29.42). G binds to the RNA and then attacks the 5′ splice site to form a phosphodiester bond with the 5′ end of

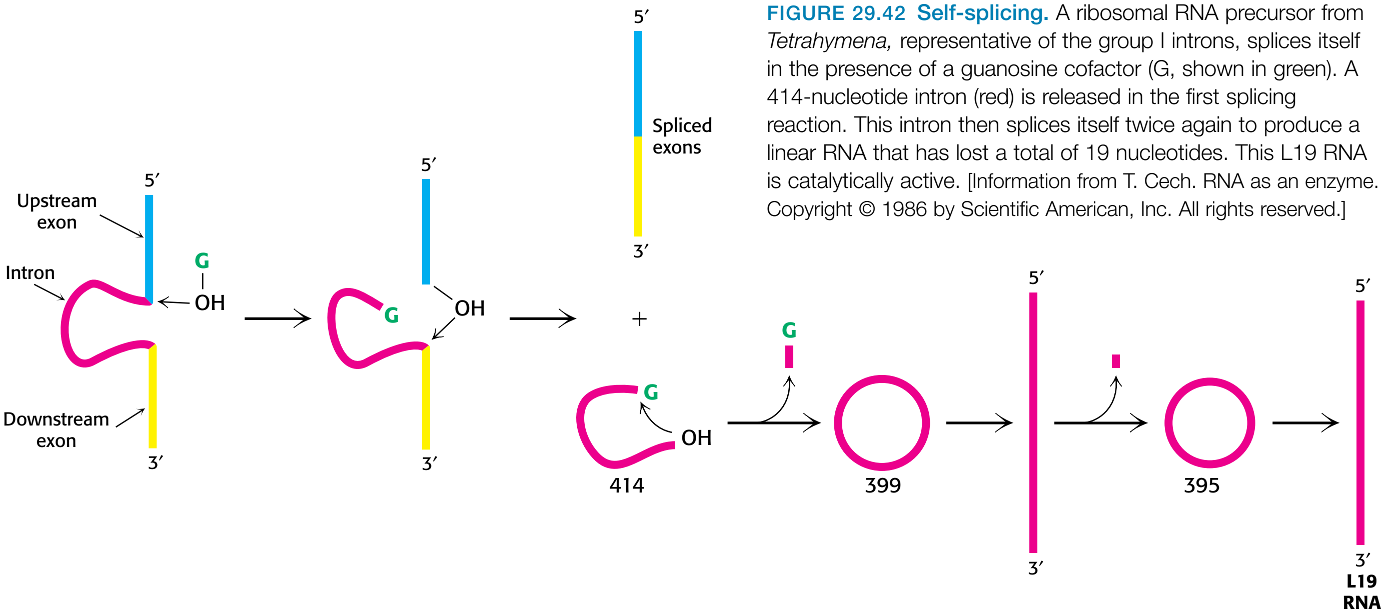


FIGURE 29.42 Self-splicing. A ribosomal RNA precursor from *Tetrahymena,* representative of the group I introns, splices itself in the presence of a guanosine cofactor (G, shown in green). A 414-nucleotide intron (red) is released in the first splicing reaction. This intron then splices itself twice again to produce a linear RNA that has lost a total of 19 nucleotides. This L19 RNA is catalytically active. [Information from T. Cech. RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]

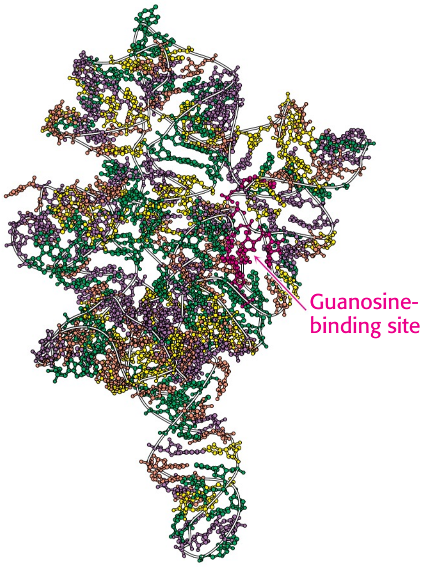


FIGURE 29.43 Structure of a self-splicing intron. The structure of a large fragment of the self-splicing intron from *Tetrahymena* reveals a complex folding pattern of helices and loops. Bases are shown in green, A; yellow, C; purple, G; and orange, U. [Drawn from 1GRZ.pdb].

the intron. This transesterification reaction generates a 3′-OH group at the end of the upstream exon. This newly attached 3′-OH group then attacks the 3′ splice site. This second transesterification reaction joins the two exons and leads to the release of the 414-nucleotide intron.

Self-splicing depends on the structural integrity of the RNA precursor. Much of the group I intron is needed for self-splicing. This molecule, like many RNAs, has a folded structure formed by many double-helical stems and loops (Figure 29.43), with a well-defined pocket for binding the guanosine. Examination of the three-dimensional structure of a catalytically active group I intron determined by x-ray crystallography reveals the coordination of magnesium ions in the active site analogous to that observed in protein enzymes such as DNA polymerase.

Analysis of the base sequence of the rRNA precursor suggested that the splice sites are aligned with the catalytic residues by base-pairing between the *internal guide sequence* (IGS) in the intron and the 5′ and 3′ exons (Figure 29.44). The IGS first brings together the guanosine cofactor and the 5′ splice site so that the 3′-OH group of G can nucleophilically attack the phosphorus atom at this splice site. The IGS then holds the downstream exon in position for attack by the newly formed 3′-OH group of the upstream exon. A phosphodiester bond is formed between the two exons, and the intron is released as a linear molecule. Like catalysis by protein enzymes, self-catalysis of bond formation and breakage in this rRNA precursor is highly specific.

The finding of enzymatic activity in the self-splicing intron and in the RNA component of RNase P has opened new areas of inquiry and changed the way in which we think about molecular evolution. As mentioned in an earlier chapter, the discovery that RNA can be a catalyst as well as an information carrier suggests that an RNA world may have existed early in the evolution of life, before the appearance of DNA and protein.

Messenger RNA precursors in the mitochondria of yeast and fungi also undergo self-splicing, as do some RNA precursors in the chloroplasts of unicellular organisms such as *Chlamydomonas*. Self-splicing reactions can be classified according to the nature of the unit that attacks the upstream splice site. Group I self-splicing is mediated by a guanosine cofactor, as in *Tetrahymena*. The attacking moiety in group II splicing is the 2′-OH group of a specific adenylate of the intron (Figure 29.45).

Group I and group II self-splicing resembles spliceosome-catalyzed splicing in two respects. First, in the initial step, a ribose hydroxyl group

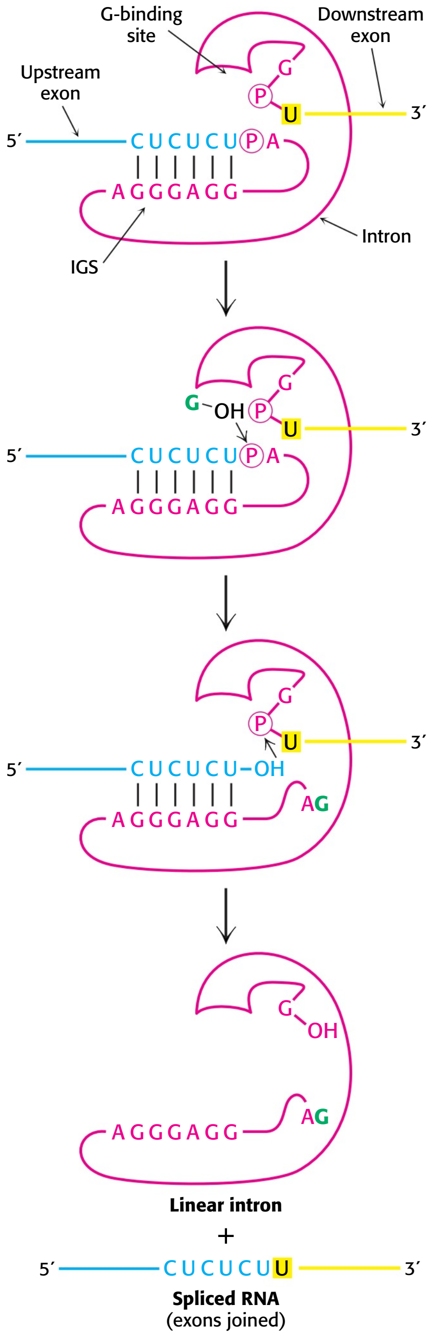


FIGURE 29.44 Self-splicing mechanism. The catalytic mechanism of the group I intron includes a series of transesterification reactions. [Information from T. Cech. RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]

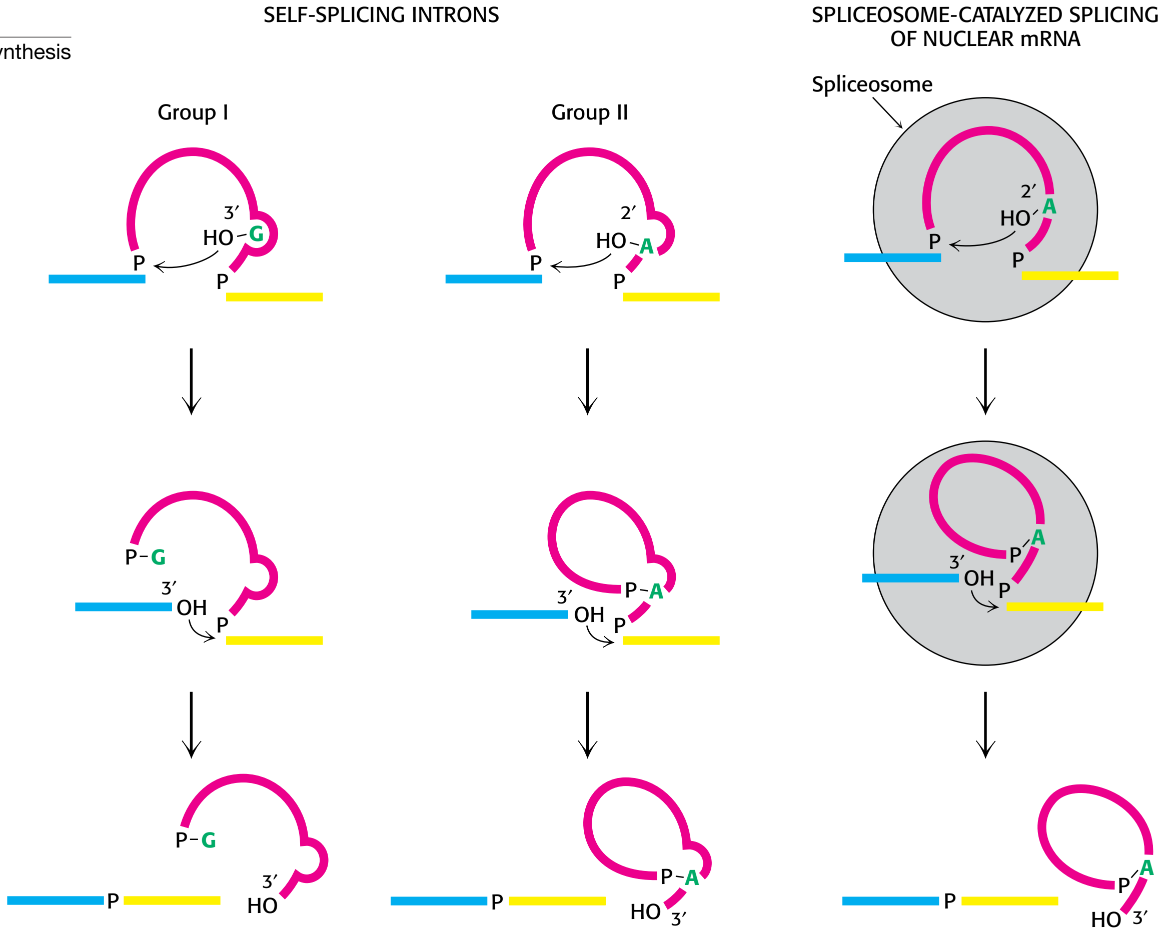


FIGURE 29.45 Comparison of splicing pathways. The exons being joined are shown in blue and yellow and the attacking unit is shown in green. The catalytic site is formed by the intron itself (red) in group I and group II splicing. In contrast, the splicing of nuclear mRNA precursors is catalyzed by snRNAs and their associated proteins in the spliceosome. [Information from P. A. Sharp, *Science* 235:766–771, 1987.]

attacks the 5′ splice site. The newly formed 3′-OH terminus of the upstream exon then attacks the 3′ splice site to form a phosphodiester bond with the downstream exon. Second, both reactions are transesterifications in which the phosphate moieties at each splice site are retained in the products. The number of phosphodiester bonds stays constant. Group II splicing is like the spliceosome-catalyzed splicing of mRNA precursors in several additional ways. The attack at the 5′ splice site is carried out by a part of the intron itself (the 2′-OH group of adenosine) rather than by an external cofactor (G). In both cases, the intron is released in the form of a lariat. Moreover, in some instances, the group II intron is transcribed in pieces that assemble through hydrogen bonding to the catalytic intron, in a manner analogous to the assembly of the snRNAs in the spliceosome.

These similarities have led to the suggestion that the spliceosome-catalyzed splicing of mRNA precursors evolved from RNA-catalyzed self-splicing. Group II splicing may well be an intermediate between group I splicing and the splicing in the nuclei of higher eukaryotes. *A major step in this transition was the transfer of catalytic power from the intron itself to other molecules.* The formation of spliceosomes gave genes a new freedom because introns were no longer constrained to provide the catalytic center for splicing. Another advantage of external catalysts for splicing is that they can be more readily regulated. However, it is important to note that similarities do not establish ancestry. The similarities between group II introns and mRNA splicing may be a result of convergent evolution. Perhaps there are only a limited number of ways to carry out efficient, specific intron excision. The determination of whether these similarities stem from ancestry or from chemistry will require expanding our understanding of RNA biochemistry.

SUMMARY

29.1 RNA Polymerases Catalyze Transcription

All cellular RNA molecules are synthesized by RNA polymerases according to instructions given by DNA templates. The activated monomer substrates are ribonucleoside triphosphates. The direction of RNA synthesis is $5' \longrightarrow 3'$, as in DNA synthesis. RNA polymerases, unlike DNA polymerases, do not need a primer.

RNA polymerase in *E. coli* is a multisubunit enzyme. The subunit composition of the ~500-kDa holoenzyme is $\alpha_2\beta\beta'\omega\sigma$ and that of the core enzyme is $\alpha_2\beta\beta'\omega$. Transcription is initiated at promoter sites consisting of two sequences, one centered near −10 and the other near −35; that is, 10 and 35 nucleotides away from the start site in the 5′ (upstream) direction. The consensus sequence of the −10 region is TATAAT. The σ subunit enables the holoenzyme to recognize promoter sites. When the growth temperature is raised, *E. coli* expresses a special σ subunit that selectively binds the distinctive promoter of heat-shock genes. RNA polymerase must unwind the template double helix for transcription to take place. Unwinding exposes some 17 bases on the template strand and sets the stage for the formation of the first phosphodiester bond. Newly synthesized RNA chains usually start with pppG or pppA. The σ subunit usually dissociates from the holoenzyme after the initiation of the new chain.

Elongation takes place at transcription bubbles that move along the DNA template at a rate of about 50 nucleotides per second. RNA polymerase occasionally backtracks, a process that can facilitate proofreading of the RNA transcript. The nascent RNA chain contains stop signals that end transcription. One stop signal is an RNA hairpin, which is followed by several U residues. A different stop signal is read by the *rho* protein, an ATPase. Some genes are regulated by riboswitches, structures that form in RNA transcripts and bind specific metabolites. In *E. coli,* precursors of transfer RNA and ribosomal RNA are cleaved and chemically modified after transcription, whereas messenger RNA is used unchanged as a template for protein synthesis.

29.2 Transcription in Eukaryotes Is Highly Regulated

RNA synthesis in eukaryotes takes place in the nucleus, whereas protein synthesis takes place in the cytoplasm. There are three types of RNA polymerase in the nucleus: RNA polymerase I makes ribosomal RNA precursors, II makes messenger RNA precursors, and III makes transfer RNA precursors. Eukaryotic promoters are complex, being composed of several different elements. Promoters for RNA polymerase II may be located on the 5′ side or the 3′ side of the start site for transcription. A common type of eukaryotic promoter consists of a TATA box centered between −30 and −100 and paired with an initiator element. Eukaryotic promoter elements are recognized by proteins called transcription factors rather than by RNA polymerase II. The saddle-shaped TATA-box-binding protein unwinds and sharply bends DNA at TATA-box sequences and serves as a focal point for the assembly of transcription complexes. The TATA-box-binding protein initiates the assembly of the active transcription complex. The activity of many promoters is greatly increased by enhancer sequences that have no promoter activity of their own. Enhancer sequences can act over distances of several kilobases, and they can be located either upstream or downstream of a gene.

29.3 The Transcription Products of Eukaryotic Polymerases Are Processed

The 5′ ends of mRNA precursors become capped and methylated in the course of transcription. A 3′ poly(A) tail is added to most mRNA precursors after the nascent chain has been cleaved by an endonuclease. RNA editing alters the nucleotide sequence of some mRNAs, such as the one for apolipoprotein B.

The splicing of mRNA precursors is carried out by spliceosomes, which consist of small nuclear ribonucleoprotein particles. Splice sites in mRNA precursors are specified by sequences at ends of introns and by branch sites near their 3′ ends. The 2′-OH group of an adenosine residue in the branch site attacks the 5′ splice site to form a lariat intermediate. The newly generated 3′-OH terminus of the upstream exon then attacks the 3′ splice site to become joined to the downstream exon. Splicing thus consists of two transesterification reactions, with the number of phosphodiester bonds remaining constant during reactions. Small nuclear RNAs in spliceosomes catalyze the splicing of mRNA precursors. In particular, U2 and U6 snRNAs form the active centers of spliceosomes.

The events in posttranscriptional processing of mRNA are controlled by the phosphorylation state of the carboxy-terminal domain, part of RNA polymerase II.

29.4 The Discovery of Catalytic RNA Was Revealing in Regard to Both Mechanism and Evolution

Some RNA molecules, such as those containing the group I intron, undergo self-splicing in the absence of protein. A self-modified version of this rRNA intron displays true catalytic activity and is thus a ribozyme. Spliceosome-catalyzed splicing may have evolved from self-splicing. The discovery of catalytic RNA has opened new vistas in our exploration of early stages of molecular evolution and the origins of life.

KEY TERMS

transcription (p. 859)
RNA polymerase (p. 860)
promoter site (p. 860)
transcription bubble (p. 862)
consensus sequence (p. 864)
sigma (σ) subunit (p. 865)
riboswitch (p. 867)
rho (ρ) protein (p. 868)
carboxy-terminal domain (CTD) (p. 872)
TATA box (p. 874)
transcription factor (p. 874)
enhancer (p. 876)
small nucleolar ribonucleoprotein (snoRNP) (p. 877)
pre-mRNA (p. 878)
5′ cap (p. 878)
poly(A) tail (p. 878)
microRNA (p. 879)
RNA editing (p. 879)
RNA splicing (p. 880)
spliceosome (p. 881)
small nuclear RNA (snRNA) (p. 882)
small nuclear ribonucleoprotein particle (snRNP) (p. 882)
alternative splicing (p. 885)
catalytic RNA (ribozyme) (p. 886)
self-splicing (p. 886)

PROBLEMS

1. *Complements.* The sequence of part of an mRNA is

5′–AUGGGGAACAGCAAGAGUGGGGCCCUGUCCAAGGAG–3′

What is the sequence of the DNA coding strand? Of the DNA template strand?

2. *Checking for errors.* Why is RNA synthesis not as carefully monitored for errors as is DNA synthesis?

3. *Speed is not of the essence.* Why is it advantageous for DNA synthesis to be more rapid than RNA synthesis?

4. *Active sites*. The overall structures of RNA polymerase and DNA polymerase are very different, yet their active sites show considerable similarities. What do the similarities suggest about the evolutionary relationship between these two important enzymes?

5. *Potent inhibitor*. Heparin inhibits transcription by binding to RNA polymerase. What properties of heparin allow it to bind so effectively to RNA polymerase?

6. *A loose cannon*. Sigma protein by itself does not bind to promoter sites. Predict the effect of a mutation enabling σ to bind to the -10 region in the absence of other subunits of RNA polymerase.

7. *Stuck sigma*. What would be the likely effect of a mutation that prevents σ from dissociating from the RNA polymerase core?

8. *Transcription time*. What is the minimum length of time required for the synthesis by *E. coli* polymerase of an mRNA encoding a 100-kDa protein?

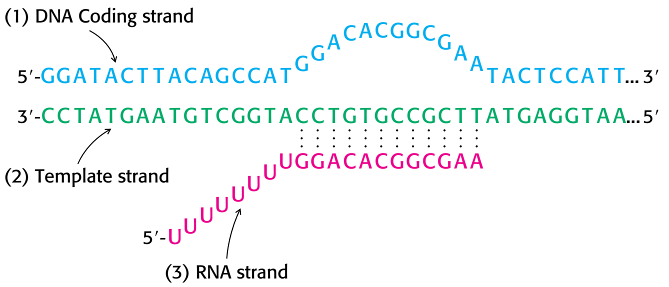
9. *Rapid search*. RNA polymerase finds promoter sites very rapidly. The observed rate constant for the binding of the RNA polymerase holoenzyme to promoter sequences is $10^{10}\ \mathrm{M}^{-1}\mathrm{s}^{-1}$. The rate constant for two macromolecules encountering each other is typically $10^{8}\ \mathrm{M}^{-1}\mathrm{s}^{-1}$. Propose an explanation for the 100-fold larger rate for a protein finding a particular site along a DNA molecule.

10. *Where to begin?* Identify the likely transcription start site in the following DNA sequence:

5′–GCCGTTGACACCGTTCGGCGATCGATCCGCTATAATGTGTGGATCCGCTT–3′
3′–CGGCAACTGTGGCAAGCCGCTAGCTAGGCGATATTACACACCTAGGCGAA–5′

11. *Between bubbles*. How far apart are transcription bubbles on *E. coli* genes that are being transcribed at a maximal rate?

12. *A revealing bubble*. Consider the synthetic RNA–DNA transcription bubble illustrated here. Refer to the coding DNA strand, the template strand, and the RNA strand as strands 1, 2, and 3, respectively.


(1) DNA Coding strand
5′-GGATACTTACAGCCATGGACACGGCGAATACTCCATT...3′
3′-CCTATGAATGTCGGTACCTGTGCCGCTTATGAGGTAA...5′
(2) Template strand
5′-UUUUUUUUGGACACGGCGAA
(3) RNA strand

(a) Suppose that strand 3 is labeled with $^{32}\mathrm{P}$ at its 5′ end and that polyacrylamide gel electrophoresis is carried out under nondenaturing conditions. Predict the autoradiographic pattern for (i) strand 3 alone, (ii) strands 1 and 3, (iii) strands 2 and 3, (iv) strands 1, 2, and 3, and (v) strands 1, 2, and 3 and core RNA polymerase.

(b) What is the likely effect of rifampicin on RNA synthesis in this system?

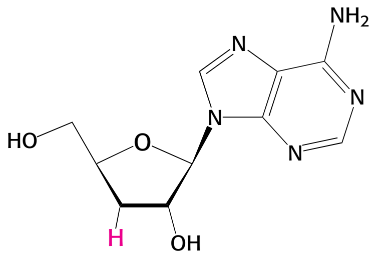
(c) Heparin blocks elongation of the RNA primer if it is added to core RNA polymerase before the onset of transcription but not if added after transcription starts. Account for this difference.

(d) Suppose that synthesis is carried out in the presence of ATP, CTP, and UTP. Compare the length of the longest product obtained with that expected when all four ribonucleoside triphosphates are present.

13. *Proofreading marks*. The major products of proofreading by RNA polymerase are dinucleotides rather than mononucleotides. Why?

14. *Abortive cycling*. Di- and trinucleotides are occasionally released from RNA polymerase at the very start of transcription, a process called abortive cycling. This process requires the restart of transcription. Suggest a plausible explanation for abortive cycling.

15. *Polymerase inhibition*. Cordycepin inhibits poly(A) synthesis at low concentrations and RNA synthesis at higher concentrations.



Cordycepin (3'-deoxyadenosine)

(a) What is the basis of inhibition by cordycepin?

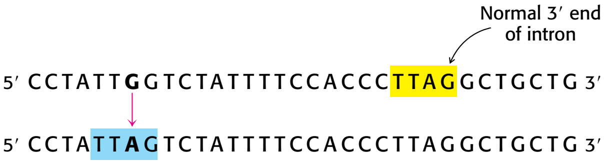
(b) Why is poly(A) synthesis more sensitive than the synthesis of other RNAs to the presence of cordycepin?

(c) Does cordycepin need to be modified to exert its effect?

16. *Alternative splicing*. A gene contains eight sites where alternative splicing is possible. Assuming that the splicing pattern at each site is independent of that at all other sites, how many splicing products are possible?

17. *Supercoiling*. Negative supercoiling of DNA favors the transcription of genes because it facilitates unwinding. However, not all promoter sites are stimulated by negative supercoiling. The promoter site for topoisomerase II itself is a noteworthy exception. Negative supercoiling decreases the rate of transcription of this gene. Propose a possible mechanism for this effect and suggest a reason why it may occur.

18. *An extra piece.* In one type of mutation leading to a form of thalassemia, the mutation of a single base (G to A) generates a new 3′ splice site (blue in the illustration below) akin to the normal one (yellow) but farther upstream.


Normal 3′ end of intron
5′ CCTATTGGTCTATTTTCCACCCTTAGGCTGCTG 3′
5′ CCTATTAGTCTATTTTCCACCCTTAGGCTGCTG 3′

What is the amino acid sequence of the extra segment of protein synthesized in a thalassemic patient having a mutation leading to aberrant splicing? The reading frame after the splice site begins with TCT.

19. *A long-tailed messenger.* Another thalassemic patient had a mutation leading to the production of an mRNA for the β chain of hemoglobin that was 900 nucleotides longer than the normal one. The poly(A) tail of this mutant mRNA was located a few nucleotides after the only AAUAAA sequence in the additional sequence. Propose a mutation that would lead to the production of this altered mRNA.

Mechanism Problem

20. *RNA editing.* Many uridine molecules are inserted into some mitochondrial mRNAs in trypanosomes. The uridine residues come from the poly(U) tail of a donor strand. Nucleoside triphosphates do not participate in this reaction. Propose a reaction mechanism that accounts for these findings. (Hint: Relate RNA editing to RNA splicing.)

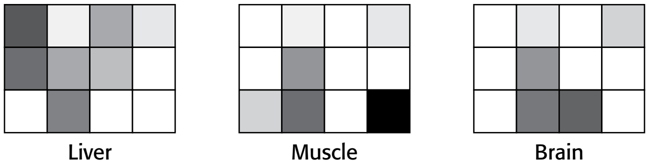
Chapter Integration Problems

21. *Proteome complexity.* What processes considered in this chapter make the proteome more complex than the genome? What processes might further enhance this complexity?

22. *Separation technique.* Suggest a means by which you could separate mRNA from the other types of RNA in a eukaryotic cell.

Data Interpretation Problems

23. *Run-off experiment.* Nuclei were isolated from brain, liver, and muscle. The nuclei were then incubated with α-[^{32}P]UTP under conditions that allow RNA synthesis, except that an inhibitor of RNA initiation was present. The radioactive RNA was isolated and annealed to various DNA sequences that had been attached to a gene chip. In the adjoining graphs, the intensity of the shading indicates roughly how much mRNA was attached to each DNA sequence.



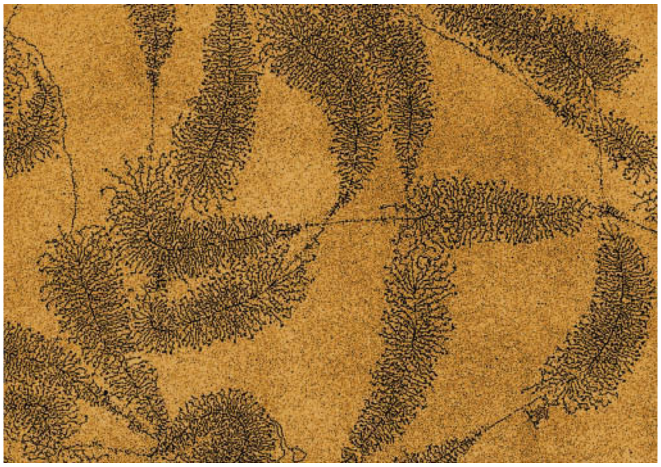
(a) Why does the intensity of hybridization differ between genes?

(b) What is the significance of the fact that some of the RNA molecules display different hybridization patterns in different tissues?

(c) Some genes are expressed in all three tissues. What would you guess is the nature of these genes?

(d) Suggest a reason why an initiation inhibitor was included in the reaction mixture.

24. *Christmas trees.* The adjoining autoradiograph depicts several bacterial genes undergoing transcription. Identify the DNA. What are the strands of increasing length? Where is the beginning of transcription? The end of transcription? On the page, what is the direction of RNA synthesis? What can you conclude about the number of enzymes participating in RNA synthesis on a given gene?

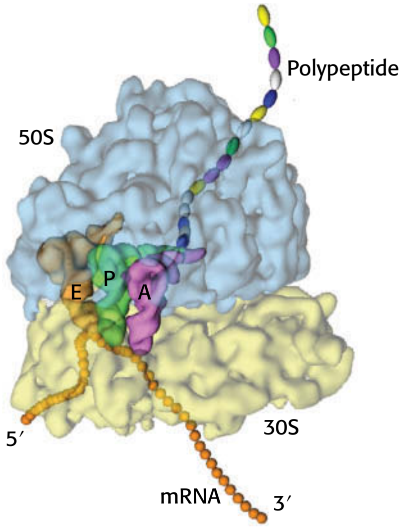


[Thomas Broker/Phototake.]

CHAPTER

30

Protein Synthesis



The ribosome, shown at the right, is a factory for the manufacture of polypeptides. Amino acids are carried into the ribosome, one at a time, connected to transfer RNA molecules. Each amino acid is joined to the growing polypeptide chain, which detaches from the ribosome only after the polypeptide has been completed. This assembly-line approach allows even very long polypeptide chains to be assembled rapidly and with impressive accuracy. [(Left) Images of Birmingham Premium/Alamy.]

The genes that encode proteins are key components of the genetic information because proteins play most of the functional roles in cells. In Chapters 28 and 29, we examined how DNA is replicated and transcribed into RNA. We now turn to the mechanism of protein synthesis, a process called *translation* because the four-letter alphabet of nucleic acids is translated into the entirely different twenty-letter alphabet of proteins. Translation is a conceptually more complex process than either replication or transcription, both of which take place within the framework of a common base-pairing language. As befits its position linking the nucleic acid and protein languages, the process of protein synthesis critically depends on both nucleic acid and protein factors. Protein synthesis takes place on *ribosomes*—enormous complexes containing three large RNA molecules and more than 50 proteins. Interestingly, *the ribosome is a ribozyme;* that is, the RNA components catalyze protein synthesis. This observation strongly supports the notion that life evolved through an RNA world, and the ribosome is a surviving inhabitant of that world.

Transfer RNA molecules (tRNAs) and messenger RNA (mRNA) also are key participants in protein synthesis. The link between amino acids and nucleic acids is first made by enzymes called aminoacyl-tRNA synthetases. By specifically linking a particular amino acid to each tRNA, these enzymes translate the genetic code.

Although RNA is paramount in the process of translation, protein factors also are required for the efficient synthesis of a protein. Protein factors participate in the initiation, elongation, and termination of protein synthesis.

OUTLINE

30.1 Protein Synthesis Requires the Translation of Nucleotide Sequences into Amino Acid Sequences

30.2 Aminoacyl Transfer RNA Synthetases Read the Genetic Code

30.3 The Ribosome Is the Site of Protein Synthesis

30.4 Eukaryotic Protein Synthesis Differs from Bacterial Protein Synthesis Primarily in Translation Initiation

30.5 A Variety of Antibiotics and Toxins Can Inhibit Protein Synthesis

30.6 Ribosomes Bound to the Endoplasmic Reticulum Manufacture Secretory and Membrane Proteins

This chapter focuses primarily on protein synthesis in bacteria because it illustrates many general principles and is well understood. Some distinctive features of protein synthesis in eukaryotes also are presented.

30.1 Protein Synthesis Requires the Translation of Nucleotide Sequences into Amino Acid Sequences

The basics of protein synthesis are the same across all kingdoms of life—evidence that the protein-synthesis system arose very early in evolution. An mRNA is decoded, or read, in the 5′-to-3′ direction, one codon at a time, and the corresponding protein is synthesized in the amino-to-carboxyl direction by the sequential addition of amino acids to the carboxyl end of the growing peptide chain (Figure 30.1). The amino acids arrive at the growing chain in activated form as aminoacyl-tRNAs, created by joining the carboxyl group of an amino acid to the 3′ end of a tRNA molecule. The linking of an amino acid to its corresponding tRNA is catalyzed by an *aminoacyl-tRNA synthetase.* For each amino acid, there is usually one activating enzyme and at least one kind of tRNA.

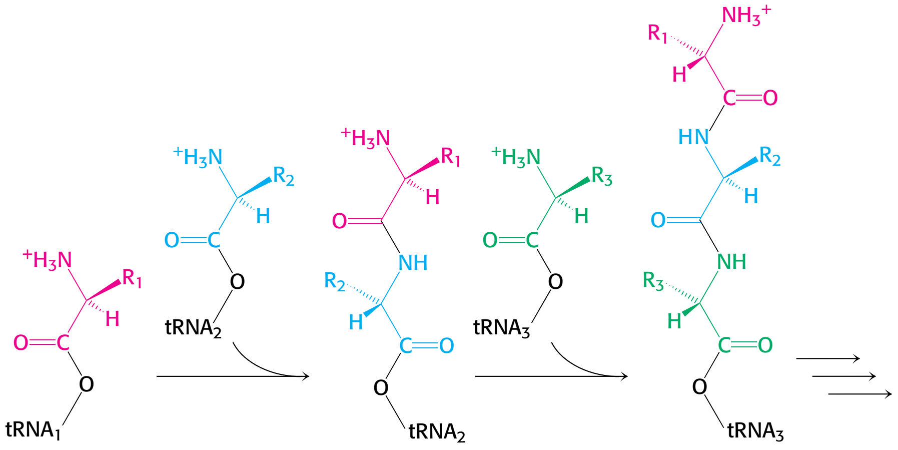


FIGURE 30.1 Polypeptide-chain growth. Proteins are synthesized by the successive addition of amino acids to the carboxyl terminus.

The synthesis of long proteins requires a low error frequency

The process of transcription is analogous to copying, word for word, a page of a book. There is no change of alphabet or vocabulary; so the likelihood of a change in meaning is small. Translating the base sequence of an mRNA molecule into a sequence of amino acids is analogous to translating the page of a book into another language. Translation is a complex process, entailing many steps and dozens of molecules. The potential for error exists at each step. The complexity of translation creates a conflict between two requirements: the process must be both accurate and fast enough to meet a cell's needs. In *E. coli,* translation can take place at a rate of 20 amino acids per second, a truly impressive speed considering the complexity of the process.

How accurate must protein synthesis be? Let us consider error rates. The probability of forming a protein with no errors depends on the number of amino acid residues and on the frequency (ε) of insertion of a wrong amino acid. As Table 30.1 shows, an error frequency of 10^{-2} is intolerable, even for quite small proteins. An ε value of 10^{-3} usually leads to the error-free synthesis of a 300-residue protein (~33 kDa) but not of a 1000-residue protein (~110 kDa). Thus, the error frequency must not exceed approximately

TABLE 30.1 Accuracy of protein synthesis

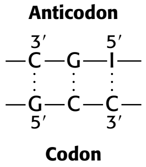
Frequency of inserting an incorrect amino acid	Probability of synthesizing an error-free protein		
	Number of amino acid residues		
	100	300	1000
10^{-2}	0.366	0.049	0.000
10^{-3}	0.905	0.741	0.368
10^{-4}	0.990	0.970	0.905
10^{-5}	0.999	0.997	0.990

Note: The probability p of forming a protein with no errors depends on n, the number of amino acids, and ε, the frequency of insertion of a wrong amino acid: $p = (1 - \varepsilon)^n$.

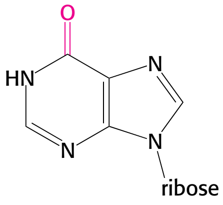
10^{-4} to produce the larger proteins effectively. Lower error frequencies are conceivable; however, except for the largest proteins, they will not dramatically increase the percentage of proteins with accurate sequences. In addition, such lower error rates are likely to be possible only by a reduction in the rate of protein synthesis because additional time for proofreading is required. *In fact, the observed values of* ε *are close to* 10^{-4}. An error frequency of about 10^{-4} per amino acid residue was selected in the course of evolution to accurately produce proteins consisting of as many as 1000 amino acids while maintaining a remarkably rapid rate for protein synthesis.

Transfer RNA molecules have a common design

The fidelity of protein synthesis requires accurate recognition of three-base *codons* on messenger RNA. Recall that the genetic code relates each amino acid to a three-letter codon (Section 4.6). An amino acid cannot itself recognize a codon. Consequently, an amino acid is attached to a specific tRNA molecule that recognizes the codon by Watson–Crick base-pairing. *Transfer RNA serves as the adapter molecule that binds to a specific codon and brings with it an amino acid for incorporation into the polypeptide chain.*



Consider yeast alanyl-tRNA, so called because it will carry the amino acid alanine. This adapter molecule is a single chain of 76 ribonucleotides (Figure 30.2). The 5′ terminus is phosphorylated (pG), whereas the 3′ terminus has a free hydroxyl group. The *amino acid-attachment site* is the 3′-hydroxyl group of the adenosine residue at the 3′ terminus of the molecule. The sequence 5′-IGC-3′ in the middle of the molecule is the *anticodon,* where I is the purine base inosine. It is complementary to 5′-GCC-3′, one of the codons for alanine.



Inosine

Thousands of tRNA sequences are known. The striking finding is that all of them can be arranged in a cloverleaf pattern in which about half the residues are base-paired (Figure 30.3). Hence, *tRNA molecules have many common structural features.* This finding is not unexpected, because all tRNA molecules must be able to interact in nearly the same way with the ribosomes, mRNAs, and protein factors that participate in translation.

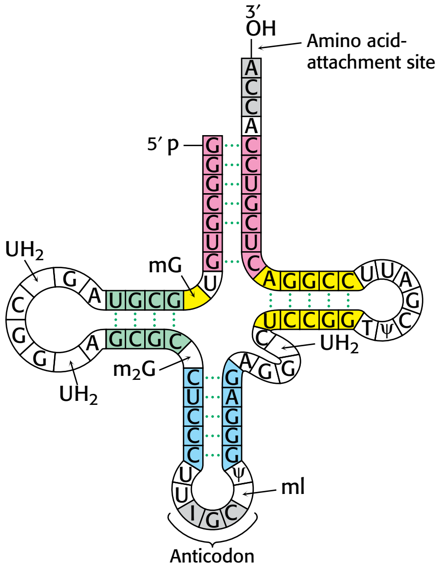


FIGURE 30.2 Alanyl-tRNA sequence. The base sequence of yeast alanyl-tRNA and the deduced cloverleaf secondary structure are shown. Modified nucleosides are abbreviated as follows: methylinosine (mI), dihydrouridine (UH_2), ribothymidine (T), pseudouridine (ψ), methylguanosine (mG), and (dimethylguanosine (m_2G). Inosine (I), another modified nucleoside, is part of the anticodon.

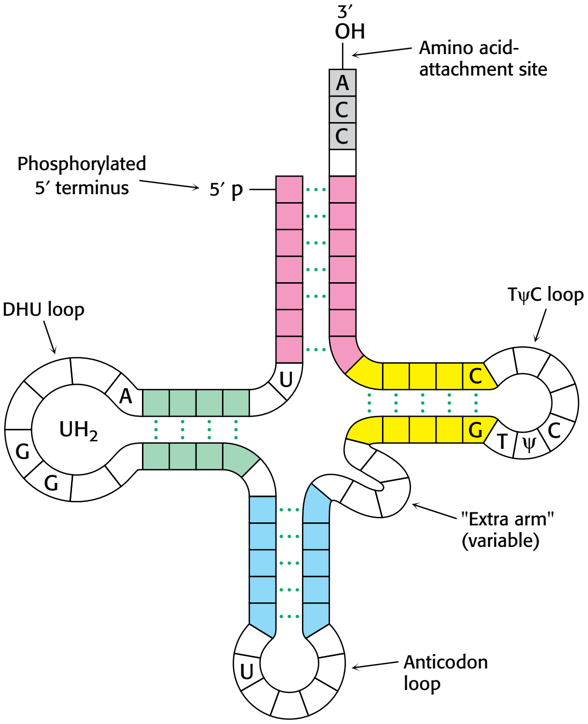


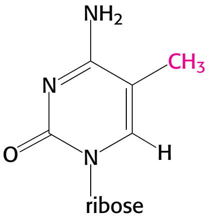
FIGURE 30.3 General structure of tRNA molecules. Comparison of the base sequences of many tRNAs reveals a number of conserved features.

All known transfer RNA molecules have the following features:

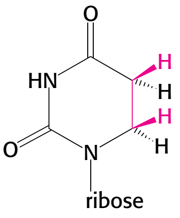
1. Each is a single chain containing between *73 and 93 ribonucleotides* (~25 kDa).

2. The molecule is L-shaped (Figure 30.4).

3. They contain *many unusual bases,* typically between 7 and 15 per molecule. Some of these bases are methylated or dimethylated derivatives of A, U, C, and G formed by enzymatic modification of a precursor tRNA. Some methylations prevent the formation of certain base pairs, thereby rendering such bases accessible for interactions with other bases. In addition, methylation imparts a hydrophobic character to some regions of tRNAs, which may be important for their interaction with synthetases and ribosomal proteins. Other modifications alter codon recognition, as will be described shortly.



5-Methylcytidine (mC)



Dihydrouridine (UH_2)

4. About half the nucleotides in tRNAs are base-paired to form double helices. The four helical regions are arranged to form two apparently continuous segments of double helix. These segments are like A-form DNA, as expected for an RNA helix (Section 4.2). One helix, containing the 5′ and 3′ ends, runs horizontally in the model shown in Figure 30.5. The other helix, which contains the anticodon and runs vertically in Figure 30.5, forms the other arm of the L.

Five groups of bases are not base-paired in this way: the 3′ *CCA terminal region,* which is part of a region called the *acceptor stem;* the *TψC loop,* which acquired its name from the sequence ribothymine-pseudouracil-cytosine; the *"extra arm,"* which contains a variable number of residues; *the DHU loop,* which contains several *di*hydro*u*racil residues; and the *anticodon loop.* Most of the bases in the nonhelical regions participate in hydrogen-bonding interactions, even if the interactions are not like those in Watson–Crick base pairs. The structural diversity generated by this combination of helices and loops containing modified bases ensures that the tRNAs can be uniquely distinguished, though they are structurally similar overall.

5. The 5′ end of a tRNA is phosphorylated. The 5′ terminal residue is usually pG.

6. An activated amino acid is attached to a hydroxyl group of the adenosine residue in the amino acid-attachment site, located at the end of the 3′ CCA

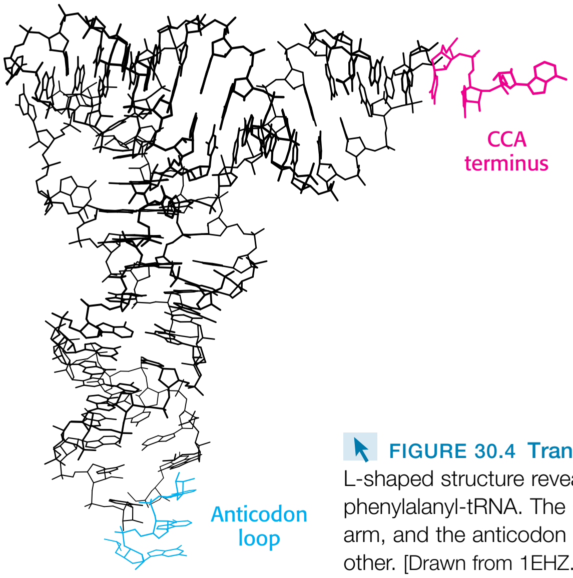


FIGURE 30.4 Transfer RNA structure. *Notice* the L-shaped structure revealed by this skeletal model of yeast phenylalanyl-tRNA. The CCA region is at the end of one arm, and the anticodon loop is at the end of the other. [Drawn from 1EHZ.pdb.]

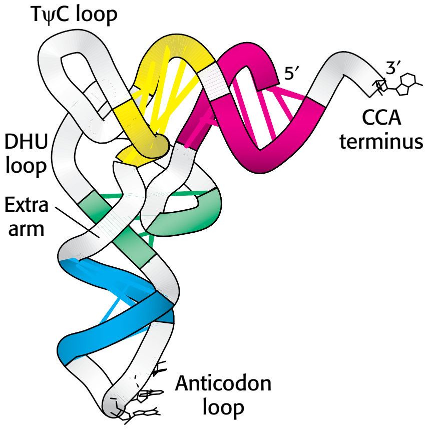


FIGURE 30.5 Helix stacking in tRNA. The four double-stranded regions of the tRNA (Figure 30.3) stack to form an L-shaped structure. [Drawn from 1EHZ.pdb.]

component of the acceptor stem (Figure 30.6). This single-stranded region can change conformation in the course of amino acid activation and protein synthesis.

7. The anticodon loop, which is present in a loop near the center of the sequence, is at the other end of the L, making accessible the three bases that constitute the anticodon.

Thus, the architecture of the tRNA molecule is well suited to its role as adaptor: the anticodon is available to interact with an appropriate codon on mRNA while the end that is linked to an activated amino acid is well positioned to participate in peptide-bond formation.

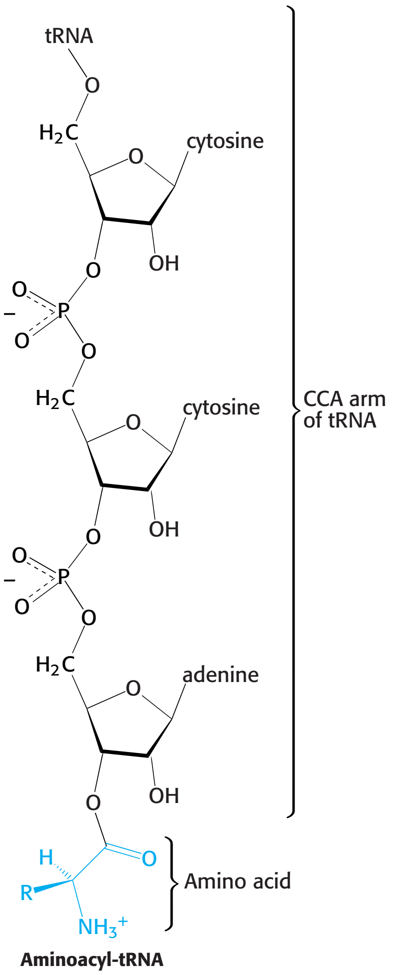
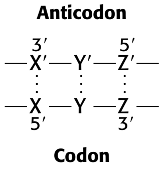


FIGURE 30.6 Aminoacyl-tRNA. Amino acids are coupled to tRNAs through ester linkages to either the 2′- or the 3′-hydroxyl group of the 3′-adenosine residue. A linkage to the 3′-hydroxyl group is shown.

Some transfer RNA molecules recognize more than one codon because of wobble in base-pairing

What are the rules that govern the recognition of a codon by the anticodon of a tRNA? A simple hypothesis is that each of the bases of the codon forms a Watson–Crick type of base pair with a complementary base on the anticodon of the tRNA. The codon and anticodon would then be lined up in an antiparallel fashion. In the diagram in the margin, the prime denotes the complementary base. Thus X and X′ would be either A and U (or U and A) or G and C (or C and G). According to this model, a particular anticodon can recognize only one codon.

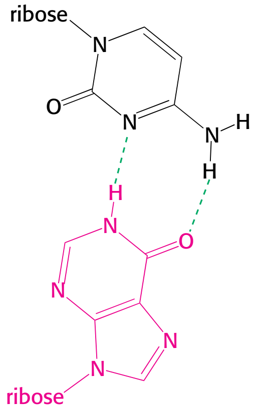


The facts are otherwise. As found experimentally, *some tRNA molecules can recognize more than one codon*. For example, the yeast alanyl-tRNA binds to *three* codons: GCU, GCC, and GCA. The first two bases of these codons are the same, whereas the third is different. Could it be that recognition of the third base of a codon is sometimes less discriminating than recognition of the other two? The pattern of degeneracy of the genetic code indicates that it might be so. XYU and XYC always encode the same amino acid; XYA and XYG usually do. These data suggest that the steric criteria might be less stringent for pairing of the third base than for the other two. In other words, there is some steric freedom (“wobble”) in the pairing of the third base of the codon.

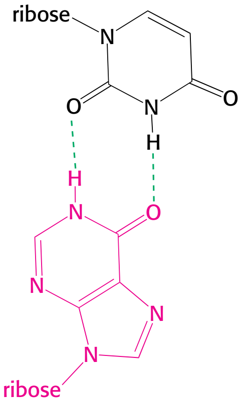
The *wobble hypothesis* is now firmly established (Table 30.2). The anticodons of tRNAs of known sequence bind to the codons predicted by this hypothesis. For example, the anticodon of yeast alanyl-tRNA is IGC. This tRNA recognizes the codons GCU, GCC, and GCA. Recall that, by convention, nucleotide sequences are written in the 5′ → 3′ direction unless otherwise noted. Hence, I (the 5′ base of this anticodon) pairs with U, C, or A (the 3′ base of the codon), as predicted.

TABLE 30.2 Allowed pairings at the third base of the codon according to the wobble hypothesis

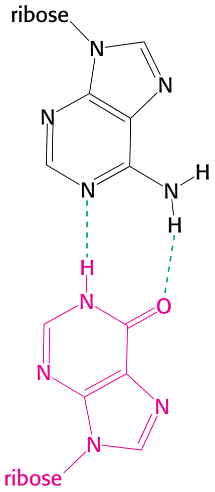
First base of anticodon	Third base of codon
C	G
A	U
U	A or G
G	U or C
I	U, C, or A



Inosine–cytidine base pair



Inosine–uridine base pair



Inosine–adenosine base pair

Two generalizations concerning the codon–anticodon interaction can be made:

1. The first two bases of a codon pair in the standard way. Recognition is precise. Hence, *codons that differ in either of their first two bases must be recognized by different tRNAs*. For example, both UUA and CUA encode leucine but are read by different tRNAs.

2. The first base of an anticodon determines whether a particular tRNA molecule reads one, two, or three kinds of codons: C or A (one codon), U or G (two codons), or I (three codons). Thus, *part of the degeneracy of the genetic code arises from imprecision (wobble) in the pairing of the third base of the codon with the first base of the anticodon*. We see here a strong reason for the frequent appearance of inosine, one of the unusual nucleosides, in anticodons. *Inosine maximizes the number of codons that can be read by a particular tRNA molecule*. The inosine bases in tRNA are formed by the deamination of adenosine after the synthesis of the primary transcript.

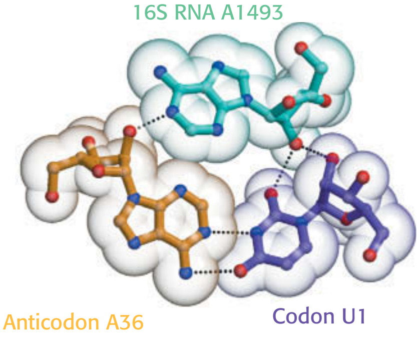


FIGURE 30.7 16S rRNA monitors base-pairing between the codon and the anticodon. Adenine 1493, one of three universally conserved bases in 16S rRNA, forms hydrogen bonds with the bases in both the codon and the anticodon only if the codon and anticodon are correctly paired. [Information from J. M. Ogle and V. Ramakrishnan, *Annu. Rev. Biochem.* 74: 129–177, 2005, Fig. 2a.]

Why is wobble tolerated in the third position of the codon but not in the first two? This question is answered by considering the interaction of the tRNA with the ribosome. As we will see, ribosomes are huge RNA–protein complexes consisting of two subunits, the 30S and 50S subunits. The 30S subunit has an RNA molecule, the 16S rRNA, that has three universally conserved bases—adenine 1492, adenine 1493, and guanine 530—that form hydrogen bonds on the minor-groove side but only with correctly formed base pairs of the codon–anticodon duplex (Figure 30.7). These interactions check whether Watson–Crick base pairs are present in the first two positions of the codon–anticodon duplex, but not the third position, so more-varied base pairs are tolerated. *Thus, the ribosome plays an active role in decoding the codon–anticodon interactions*.

30.2 Aminoacyl Transfer RNA Synthetases Read the Genetic Code

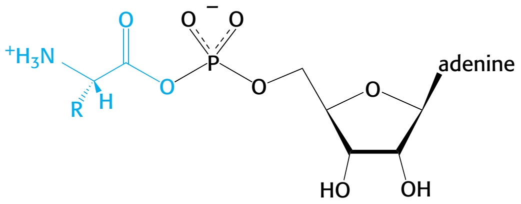
Before codon and anticodon meet, the amino acids required for protein synthesis must first be attached to specific tRNA molecules, linkages that are crucial for two reasons. *First, the attachment of a given amino acid to a particular tRNA establishes the genetic code*. When an amino acid has been linked to a tRNA, it will be incorporated into a growing polypeptide chain at a position dictated by the anticodon of the tRNA. *Second, because the formation of a peptide bond between free amino acids is not thermodynamically favorable, the amino acid must first be activated in order for protein synthesis to proceed. The activated intermediates in protein synthesis are amino acid esters,* in which the carboxyl group of an amino acid is linked to either the 2′- or the 3′-hydroxyl group of the ribose unit at the 3′ end of tRNA. An amino acid ester of tRNA is called an *aminoacyl-tRNA* or sometimes a *charged tRNA* (Figure 30.6). For a specific amino acid attached to its cognate tRNA—for instance, threonine—the charged tRNA is designated Thr-tRNA$^{\text{Thr}}$.

Amino acids are first activated by adenylation

The activation reaction is catalyzed by specific *aminoacyl-tRNA synthetases*. The first step is the formation of an *aminoacyl adenylate* from an amino acid and ATP.

$$\text{Amino acid} + \text{ATP} \rightleftharpoons \text{aminoacyl-AMP} + \text{PP}_i$$

This activated species is a mixed anhydride in which the carboxyl group of the amino acid is linked to the phosphoryl group of AMP; hence, it is also known as *aminoacyl-AMP*.



Aminoacyl adenylate

The next step is the transfer of the aminoacyl group of aminoacyl-AMP to a particular tRNA molecule to form *aminoacyl-tRNA*.

$$\text{Aminoacyl-AMP} + \text{tRNA} \rightleftharpoons \text{aminoacyl-tRNA} + \text{AMP}$$

The sum of these activation and transfer steps is

$$\text{Amino acid} + \text{ATP} + \text{tRNA} \rightleftharpoons \text{aminoacyl-tRNA} + \text{AMP} + \text{PP}_i$$

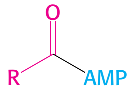
The $\Delta G^{\circ\prime}$ of this reaction is close to 0, because the free energy of hydrolysis of the ester bond of aminoacyl-tRNA is similar to that for the hydrolysis of ATP to AMP and PP_i. As we have seen many times, the reaction is driven by the hydrolysis of pyrophosphate. The sum of these three reactions is highly exergonic:

$$\text{Amino acid} + \text{ATP} + \text{tRNA} + H_2O \longrightarrow \text{aminoacyl-tRNA} + \text{AMP} + 2\,P_i$$

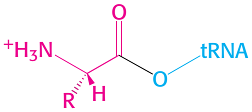
Thus, *the equivalent of two molecules of ATP is consumed in the synthesis of each aminoacyl-tRNA*. One of them is consumed in forming the ester linkage of aminoacyl-tRNA, whereas the other is consumed in driving the reaction forward.

The activation and transfer steps for a particular amino acid are catalyzed by the same aminoacyl-tRNA synthetase. Indeed, *the aminoacyl-AMP intermediate does not dissociate from the synthetase*. Rather, it is tightly bound to the active site of the enzyme by noncovalent interactions.

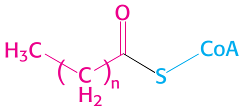
We have already encountered an acyl adenylate intermediate in fatty acid activation (Section 22.2). The major difference between these reactions is that the acceptor of the acyl group is CoA in fatty acid activation and tRNA in amino acid activation. The energetics of these biosyntheses are very similar: both are made irreversible by the hydrolysis of pyrophosphate.



Acyl adenylate intermediate



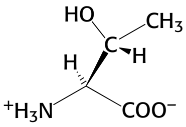
Aminoacyl-tRNA



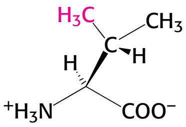
Fatty acyl CoA

Aminoacyl-tRNA synthetases have highly discriminating amino acid activation sites

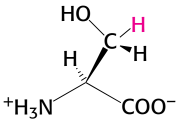
Each aminoacyl-tRNA synthetase is highly specific for a given amino acid. Indeed, a synthetase will incorporate the incorrect amino acid only once in 10^4 or 10^5 reactions. How is this level of specificity achieved? Each aminoacyl-tRNA synthetase takes advantage of the properties of its amino acid substrate. Let us consider the challenge faced by threonyl-tRNA synthetase. Threonine is especially similar to two other amino acids—namely, valine and serine. Valine has almost exactly the same shape as that of threonine, except that valine has a methyl group in place of a hydroxyl group. Serine has a hydroxyl group, as does threonine, but lacks the methyl group. How can the threonyl-tRNA synthetase avoid coupling these incorrect amino acids to threonyl-tRNA?



Threonine



Valine



Serine

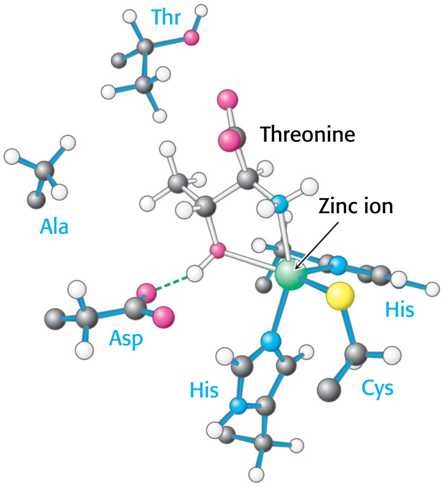


FIGURE 30.8 Active site of threonyl-tRNA synthetase. *Notice* that the amino acid-binding site includes a zinc ion (green ball) that binds threonine through its amino and hydroxyl groups.

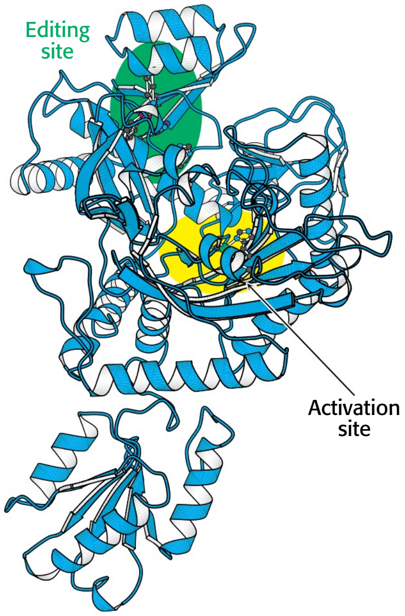


FIGURE 30.9 Editing site. Mutagenesis studies revealed the position of the editing site (shown in green) in threonyl-tRNA synthetase. The activation site is shown in yellow. Only one subunit of the dimeric enzyme is shown here and in subsequent illustrations. [Drawn from 1QF6.pdb.]

The structure of the amino acid-binding site of threonyl-tRNA synthetase reveals how valine is avoided (Figure 30.8). The synthetase contains a zinc ion, bound to the enzyme by two histidine residues and one cysteine residue. The remaining coordination sites are available for substrate binding. Threonine binds to the zinc ion through its amino group and its side-chain hydroxyl group. The side-chain hydroxyl group is further recognized by an aspartate residue that hydrogen bonds to it. The methyl group present in valine in place of this hydroxyl group cannot participate in these interactions; it is excluded from this active site and, hence, does not become adenylated and transferred to threonyl-tRNA (abbreviated tRNA$^{\mathrm{Thr}}$). The use of a zinc ion appears to be unique to threonyl-tRNA synthetase; other aminoacyl-tRNA synthetases have different strategies for recognizing their cognate amino acids. The carboxylate group of the correctly positioned threonine is available to attack the α phosphoryl group of ATP to form the aminoacyl adenylate.

The zinc site is less able to discriminate against serine because this amino acid does have a hydroxyl group that can bind to the zinc ion. Indeed, with only this mechanism available, threonyl-tRNA synthetase does mistakenly couple serine to threonyl-tRNA at a rate 10^{-2} to 10^{-3} times that for threonine. As noted on page 894, this error rate is likely to lead to many translation errors. How is a higher level of specificity achieved?

Proofreading by aminoacyl-tRNA synthetases increases the fidelity of protein synthesis

Threonyl-tRNA synthetase can be incubated with tRNA$^{\mathrm{Thr}}$ that has been covalently linked with serine (Ser-tRNA$^{\mathrm{Thr}}$); the tRNA has been "mischarged." The reaction is immediate: a rapid hydrolysis of the aminoacyl-tRNA forms serine and free tRNA. In contrast, incubation with correctly charged Thr-tRNA$^{\mathrm{Thr}}$ results in no reaction. Thus, threonyl-tRNA synthetase contains an additional functional site that hydrolyzes Ser-tRNA$^{\mathrm{Thr}}$ but not Thr-tRNA$^{\mathrm{Thr}}$. This editing site provides an opportunity for the synthetase to correct its mistakes and improve its fidelity to less than one mistake in 10^4. The results of structural and mutagenesis studies revealed that the editing site is more than 20 Å from the activation site (Figure 30.9). This editing site readily accepts and cleaves Ser-tRNA$^{\mathrm{Thr}}$ but does not cleave Thr-tRNA$^{\mathrm{Thr}}$. The discrimination of serine from threonine is easy because threonine contains an *extra* methyl group; a site that conforms to the structure of serine will sterically exclude threonine. The structure of the complex between threonyl-tRNA synthetase and its substrate reveals that the aminoacylated CCA can swing out of the activation site and into the editing site (Figure 30.10). Thus, the aminoacyl-tRNA can be edited

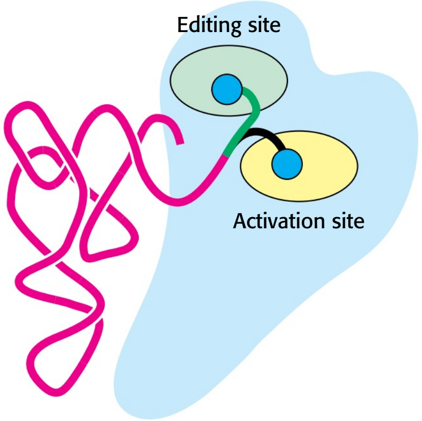


FIGURE 30.10 Editing of aminoacyl-tRNA. The flexible CCA arm of an aminoacyl-tRNA can move the amino acid between the activation site and the editing site. If the amino acid fits well into the editing site, the amino acid is removed by hydrolysis.

without dissociating from the synthetase. This proofreading depends on the conformational flexibility of a short stretch of polynucleotide sequence.

Most aminoacyl-tRNA synthetases contain editing sites in addition to activation sites. These complementary pairs of sites function as a *double sieve* to ensure very high fidelity. In general, the acylation site rejects amino acids that are *larger* than the correct one because there is insufficient room for them, whereas the hydrolytic site cleaves activated species that are *smaller* than the correct one.

A few synthetases achieve high accuracy without editing. For example, tyrosyl-tRNA synthetase has no difficulty discriminating between tyrosine and phenylalanine; the hydroxyl group on the tyrosine ring enables tyrosine to bind to the enzyme 10^4 times as strongly as phenylalanine. *Proofreading has been selected in evolution only when fidelity must be enhanced beyond what can be obtained through an initial binding interaction.*

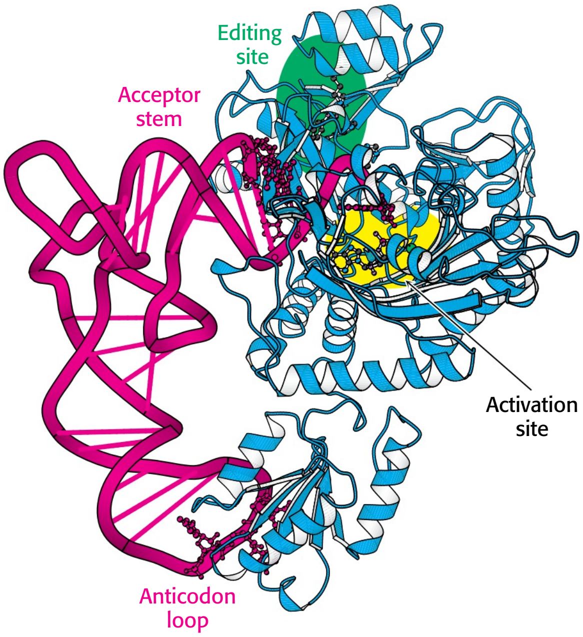


FIGURE 30.11 Threonyl-tRNA synthetase complex. The structure shows the complex between threonyl-tRNA synthetase (blue) and tRNA^{Thr} (red). *Notice* that the synthetase binds to both the acceptor stem and the anticodon loop. [Drawn from 1QF6.pdb.]

Synthetases recognize various features of transfer RNA molecules

How do synthetases choose their tRNA partners? This enormously important step is the point at which "translation" takes place—at which the correlation between the amino acid and the nucleic acid worlds is made. In a sense, aminoacyl-tRNA synthetases are the only molecules in biology that "know" the genetic code. Their precise recognition of tRNAs is as important for high-fidelity protein synthesis as is the accurate selection of amino acids. In general, tRNA recognition by the synthetase is different for each synthetase and tRNA pair. Consequently, generalities are difficult to make.

Some synthetases recognize their tRNA partners primarily on the basis of their anticodons, although they may also recognize other aspects of tRNA structure that vary among different tRNAs. The most direct evidence comes from crystallographic studies of complexes formed between synthetases and their cognate tRNAs. Consider, for example, the structure of the complex between threonyl-tRNA synthetase and tRNA^{Thr} (Figure 30.11). As expected, the CCA arm extends into the zinc-containing activation site, where it is well positioned to accept threonine from threonyl adenylate. The enzyme interacts extensively not only with the acceptor stem of the tRNA, but also with the anticodon loop. The interactions with the anticodon loop are particularly revealing. Each base within the sequence 5′-CGU-3′ of the anticodon participates in hydrogen bonds with the enzyme; those with the second two bases (G and U) appear to be more important because the synthetase interacts just as efficiently with the anticodons GGU and UGU. Although interactions between the enzyme and the anticodon are often crucial for correct recognition, Figure 30.12 shows that many aspects of tRNA molecules are recognized by synthetases. Note that many of the recognition sites are loops rich in unusual bases that can provide structural identifiers.

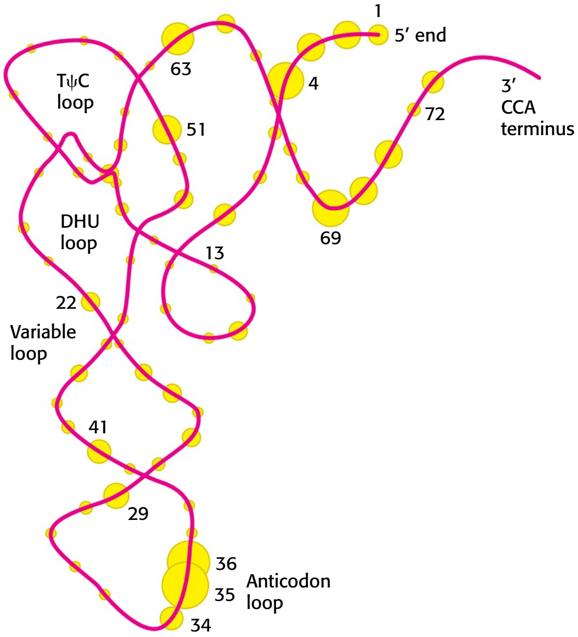


FIGURE 30.12 Recognition sites on tRNA. Circles represent nucleotides, and the sizes of the circles are proportional to the frequency with which they are used as recognition sites by aminoacyl-tRNA synthetases. The numbers indicate the positions of the nucleotides in the base sequence, beginning from the 5′ end of the tRNA molecule. [Information from M. Ibba, and D. Söll, *Annu. Rev. Biochem.* 69:617–650, 1981, p. 636.]

Aminoacyl-tRNA synthetases can be divided into two classes

At least one aminoacyl-tRNA synthetase exists for each amino acid. The diverse sizes, subunit composition, and

FIGURE 30.13 Classes of aminoacyl-tRNA synthetases. *Notice* that class I and class II synthetases recognize different faces of the tRNA molecule. The CCA arm of tRNA adopts different conformations in complexes with the two classes of synthetase. Note that the CCA arm of the tRNA is turned toward the viewer (Figures 30.4 and 30.5). [Drawn from 1EUY.pdb and 1QF6.pdb.]

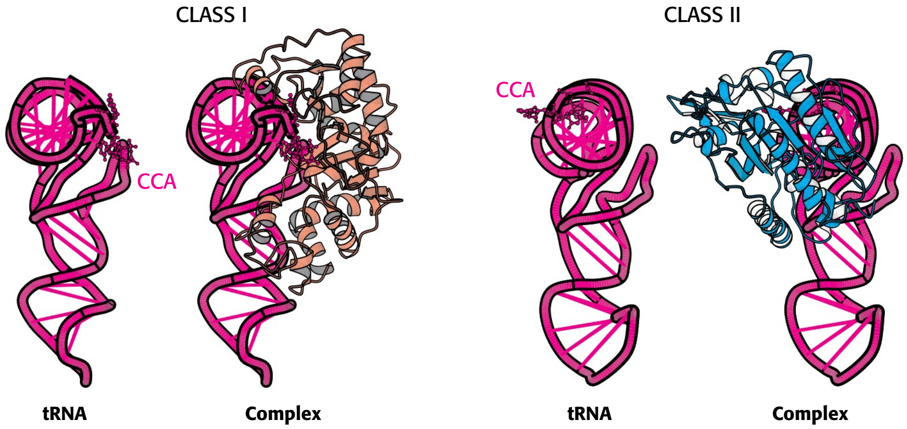


TABLE 30.3 Classification and subunit structure of aminoacyl-tRNA synthetases in *E. coli*

Class I	Class II
Arg (α)	Ala (α4)
Cys (α)	Asn (α2)
Gln (α)	Asp (α2)
Glu (α)	Gly (α2β2)
IIe (α)	His (α2)
Leu (α)	Lys (α2)
Met (α)	Phe (α2β2)
Trp (α2)	Ser (α2)
Tyr (α2)	Pro (α2)
Val (α)	Thr (α2)

sequences of these enzymes were bewildering for many years. Could it be that essentially all synthetases evolved independently? The determination of the three-dimensional structures of several synthetases followed by more-refined sequence comparisons revealed that different synthetases are, in fact, related. Specifically, synthetases fall into two classes, termed *class I* and *class II,* each of which includes enzymes specific for 10 of the 20 amino acids (Table 30.3). Intriguingly, synthetases from the two classes bind to different faces of the tRNA molecule (Figure 30.13). The CCA arm of tRNA adopts different conformations to accommodate these interactions; the arm is in the helical conformation observed for free tRNA (Figures 30.4 and 30.5) for class II enzymes and in a hairpin conformation for class I enzymes. These two classes also differ in other ways.

1. Class I enzymes acylate the 2′-hydroxyl group of the terminal adenosine of tRNA, whereas class II enzymes (except the enzyme for Phe-tRNA) acylate the 3′-hydroxyl group.
2. The two classes bind ATP in different conformations.
3. Most class I enzymes are monomeric, whereas most class II enzymes are dimeric.

Why did two distinct classes of aminoacyl-tRNA synthetases evolve? The observation that the two classes bind to distinct faces of tRNA suggests a possibility. Recognition sites on both faces of tRNA may have been required to allow the recognition of 20 different tRNAs.

30.3 The Ribosome Is the Site of Protein Synthesis

We turn now to ribosomes, the molecular machines that coordinate the interplay of aminoacyl-tRNAs, mRNA, and proteins that leads to protein synthesis. An *E. coli* ribosome is a ribonucleoprotein assembly with a mass of about 2500 kDa, a diameter of approximately 250 Å, and a sedimentation coefficient (Section 3.1) of 70S. The 20,000 ribosomes in a bacterial cell constitute nearly a fourth of its mass.

A ribosome can be dissociated into a *large subunit (50S)* and a *small subunit (30S)*. These subunits can be further split into their constituent proteins and RNAs. The 30S subunit contains 21 different proteins (referred to as S1 through S21) and a 16S RNA molecule. The 50S subunit contains 34 different proteins (L1 through L34) and two RNA molecules, a 23S and a 5S

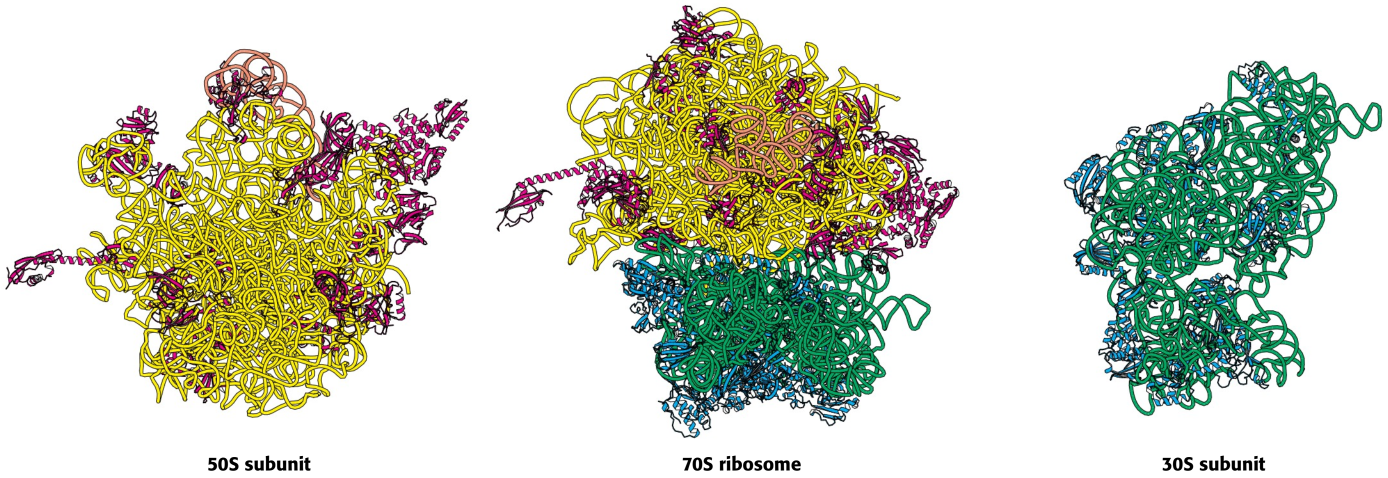


FIGURE 30.14 The ribosome at high resolution. Detailed models of the ribosome based on the results of x-ray crystallographic studies of the 70S ribosome and the 30S and 50S subunits: (left) view of the part of the 50S subunit that interacts with the 30S subunit; (center) side view of the 70S ribosome; (right) view of the part of the 30S subunit that interacts with the 50S subunit. 23S RNA is shown in yellow, 5S RNA in orange, 16S RNA in green, proteins of the 50S subunit in red, and proteins of the 30S subunit in blue. *Notice* that the interface between the 50S and the 30S subunits consists entirely of RNA. [Drawn from 1GIX. pdb and 1GIY.pdb.]

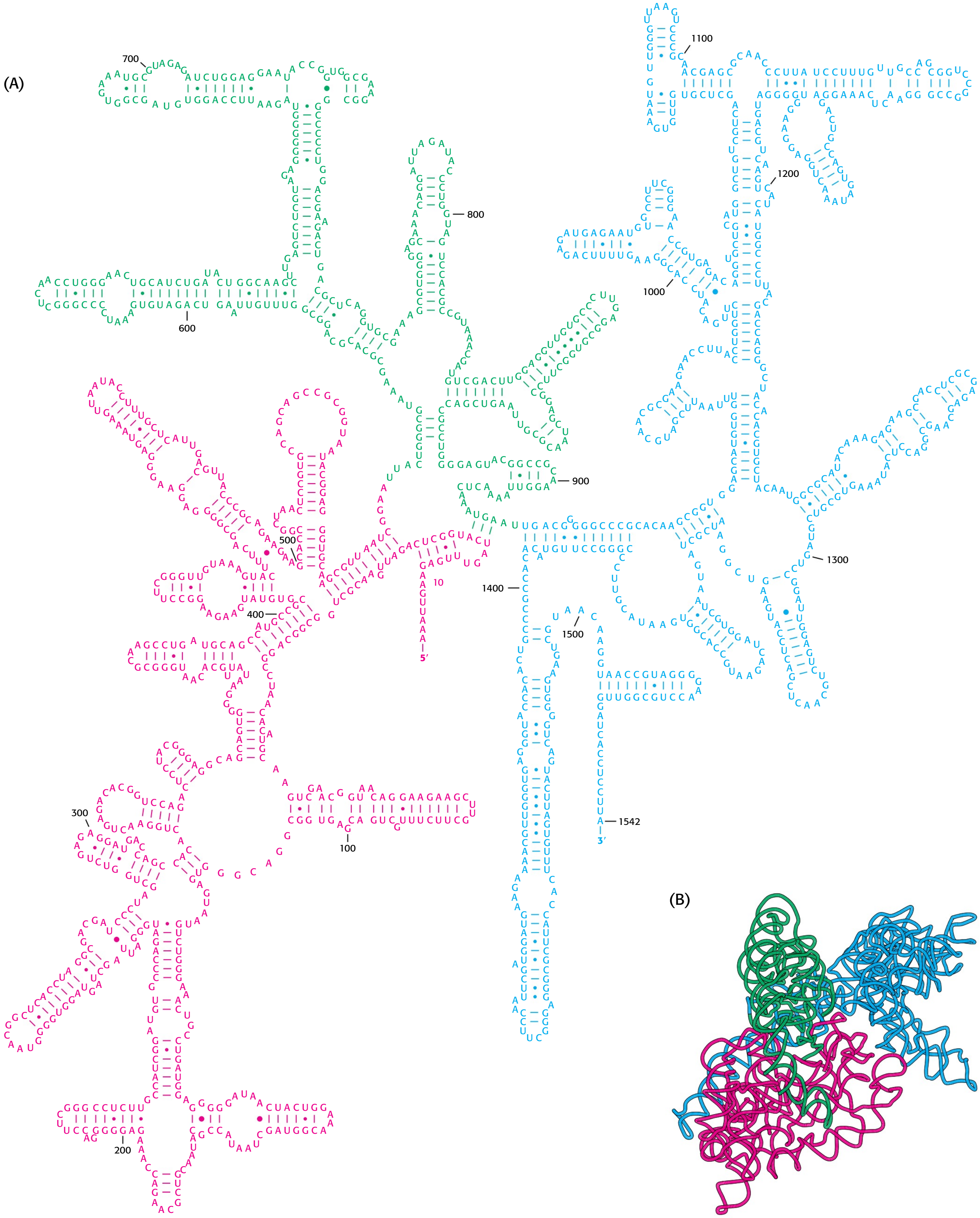
species. A ribosome contains one copy of each RNA molecule, two copies each of the L7 and L12 proteins, and one copy of each of the other proteins. The L7 protein is identical with L12 except that its amino terminus is acetylated (Section 10.3). Both the 30S and the 50S subunits can be reconstituted in vitro from their constituent proteins and RNA. *This reconstitution is an outstanding example of the principle that supramolecular complexes can form spontaneously from their macromolecular constituents*.

The structures of both the 30S and the 50S subunits as well as the complete 70S ribosome have been determined (Figure 30.14). The features of these structures are in remarkable agreement with interpretations of less-direct experimental probes. These structures provide an invaluable framework for examining the mechanism of protein synthesis.

Ribosomal RNAs (5S, 16S, and 23S rRNA) play a central role in protein synthesis

The prefix *ribo* in the name *ribosome* is apt because RNA constitutes nearly two-thirds of the mass of these large molecular assemblies. The three RNAs present—5S, 16S, and 23S—are critical for ribosomal architecture and function. They are formed by the cleavage of primary 30S transcripts and further processing. These molecules fold into structures that allow them to form internal base pairs. Their base-pairing patterns were deduced by comparing the nucleotide sequences of many species to detect conserved sequences as well as conserved base pairings. For instance, the 16S RNA of one species may have a G–C base pair, whereas another may have an A–U base pair, but the location of the base pair is the same in both molecules. Chemical modification and digestion experiments supported the structures deduced from sequence comparisons (Figure 30.15). The striking finding is that across all species, *ribosomal RNAs* (rRNAs) *are folded into defined structures that have many short duplex regions*.

For many years, ribosomal proteins were presumed to orchestrate protein synthesis and ribosomal RNAs were presumed to serve primarily as structural scaffolding. The current view is almost the reverse. The discovery of catalytic RNA (Section 29.4) made biochemists receptive to the possibility that RNA plays a much more active role in ribosomal function. The detailed structures make it clear that the key sites in the ribosome, such as those that catalyze the formation of the peptide bond and interact with mRNA and tRNA, are composed almost entirely of RNA. Contributions from the proteins are minor. The almost inescapable conclusion is that the ribosome initially consisted only of RNA and that the proteins were added later to fine-tune



(A)
5′
10
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1542
3′
(B)

FIGURE 30.15 Ribosomal RNA folding pattern. (A) The secondary structure of 16S ribosomal RNA deduced from sequence comparison and the results of chemical studies. (B) The tertiary structure of 16S RNA determined by x-ray crystallography. [(A) Courtesy of Dr. Bryn Weiser and Dr. Harry Noller; (B) drawn from 1FJG.pdb.]

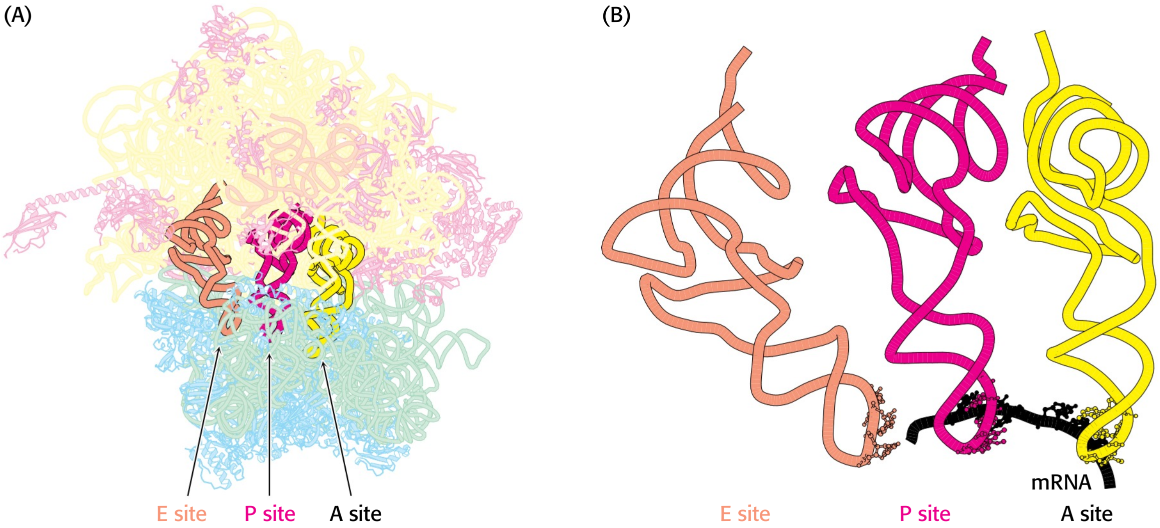


FIGURE 30.16 Transfer RNA-binding sites. (A) Three tRNA-binding sites are present on the 70S ribosome. They are called the A (for aminoacyl), P (for peptidyl), and E (for exit) sites. Each tRNA molecule contacts both the 30S and the 50S subunit. (B) The tRNA molecules in sites A and P are base-paired with mRNA. [(B) Drawn from 1JGP. pdb.]

its functional properties. This conclusion has the pleasing consequence of dodging a "chicken and egg" question: How can complex proteins be synthesized if complex proteins are required for protein synthesis?

Ribosomes have three tRNA-binding sites that bridge the 30S and 50S subunits

Three tRNA-binding sites in ribosomes are arranged to allow the formation of peptide bonds between amino acids encoded by the codons on mRNA (Figure 30.16). The mRNA fragment being translated at a given moment is bound within the 30S subunit. Each of the tRNA molecules is in contact with both the 30S subunit and the 50S subunit. At the 30S end, two of the three tRNA molecules are bound to the mRNA through anticodon–codon base pairs. These binding sites are called the A site (for *a*minoacyl) and the P site (for *p*eptidyl). The third tRNA molecule is bound to an adjacent site called the E site (for *e*xit).

The other end of each tRNA molecule, the end without the anticodon, interacts with the 50S subunit. The acceptor stems of the tRNA molecules occupying the A site and the P site converge at a site where a peptide bond is formed. A tunnel connects this site to the back of the ribosome, through which the polypeptide chain passes during synthesis (Figure 30.17).

The start signal is usually AUG preceded by several bases that pair with 16S rRNA

How does protein synthesis start? The simplest possibility would be for the first 3 nucleotides of each mRNA to serve as the first codon; no special start signal would then be needed. However, experiments show that translation in bacteria does not begin immediately at the 5′ terminus of mRNA. Indeed, the first translated codon is nearly always more than 25 nucleotides away from the 5′ end. Furthermore, in bacteria, many mRNA molecules are *polycistronic*—that is, they encode two or more polypeptide chains. For example, a single mRNA molecule about 7000-nucleotides long specifies five enzymes in the biosynthetic pathway for tryptophan in *E.coli*. Each of these five proteins has its own start and stop signals on the mRNA. In fact, *all known mRNA molecules contain signals that define the beginning and end of each encoded polypeptide chain*.

A clue to the mechanism of initiation was the finding that nearly half the amino-terminal residues of proteins in *E. coli* are methionine. In fact, the

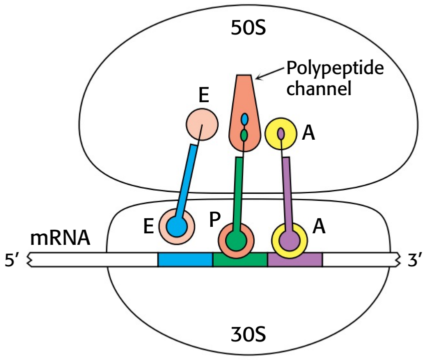


FIGURE 30.17 An active ribosome. This schematic representation shows the relations among the key components of the translation machinery.

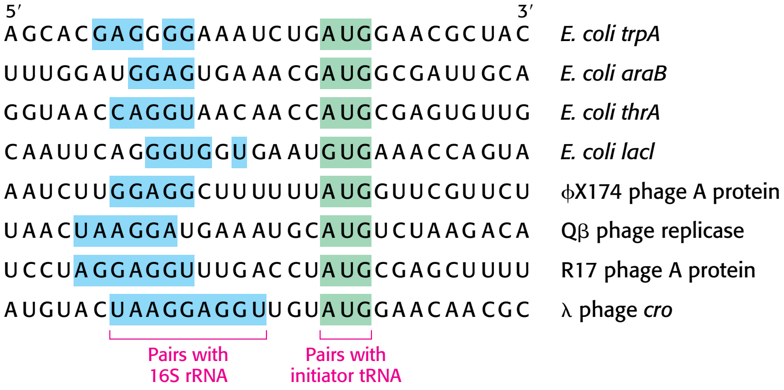

5′
3′
AGCACGAGGGGAAAUCUGAUGGAACGCUAC E. coli trpA
UUUGGAUGGAGUGAAACGAUGGCGAUUGCA E. coli araB
GGUAACCAGGUAACAACCAUGCGAGUGUUG E. coli thrA
CAAUUCAGGGUGGUGAAUGUGAAACCAGUA E. coli lacI
AAUCUUGGAGGCUUUUUUAUGGUUCGUUCU ϕX174 phage A protein
UAACUAAGGAUGAAAUGCAUGUCUAAGACA Qβ phage replicase
UCCUAGGAGGUUUGACCUAUGCGAGCUUUU R17 phage A protein
AUGUACUAAGGAGGUUGUAUGGAACAACGC λ phage cro
Pairs with 16S rRNA
Pairs with initiator tRNA

FIGURE 30.18 Initiation sites. Sequences of mRNA initiation sites for protein synthesis in some bacterial and viral mRNA molecules. Comparison of these sequences reveals some recurring features.

initiating codon in mRNA is AUG (methionine) or, less frequently, GUG (valine) or, rarely UUG (leucine). What additional signals are necessary to specify a translation start site? The first step toward answering this question was the isolation of initiator regions from a number of mRNAs. This isolation was accomplished by using ribonuclease to digest mRNA–ribosome complexes (formed under conditions in which protein synthesis could begin but elongation could not take place). As expected, each initiator region usually displays an AUG codon (Figure 30.18). In addition, each initiator region contains a purine-rich sequence centered about 10 nucleotides on the 5′ side of the initiator codon.

The role of this purine-rich region, called the *Shine–Dalgarno sequence*, (named after John Shine and Lynn Dalgarno, who first described the sequence), became evident when the sequence of 16S rRNA was elucidated. The 3′ end of this rRNA component of the 30S subunit contains a sequence of several bases that is complementary to the purine-rich region in the initiator sites of mRNA. Mutagenesis of the CCUCC sequence near the 3′ end of 16S rRNA to ACACA markedly interferes with the recognition of start sites in mRNA. This result and other evidence show that the initiator region of mRNA binds very near the 3′ end of the 16S rRNA. The number of base pairs linking mRNA and 16S rRNA ranges from three to nine. Thus, *two kinds of interactions determine where protein synthesis starts: (1) the pairing of mRNA bases with the 3′ end of 16S rRNA and (2) the pairing of the initiator codon on mRNA with the anticodon of an initiator tRNA molecule.*

Bacterial protein synthesis is initiated by formylmethionyl transfer RNA

As stated earlier, methionine is the first amino acid in many *E. coli* proteins. However, the methionine residue found at the amino-terminal end of *E. coli* proteins is usually modified. In fact, *protein synthesis in bacteria starts with the modified amino acid N-formylmethionine* (fMet). A special tRNA brings formylmethionine to the ribosome to initiate protein synthesis. This *initiator tRNA* (abbreviated as $tRNA_f$) differs from the tRNA that inserts methionine in internal positions (abbreviated as $tRNA_m$). The subscript "f" indicates that methionine attached to the initiator tRNA can be formylated, whereas it cannot be formylated when attached to $tRNA_m$. Although virtually all proteins synthesized in *E. coli* begin with formylmethionine, in approximately one-half of the proteins, *N*-formylmethionine is removed when the nascent chain is 10 amino acids long.

Methionine is linked to these two kinds of tRNAs by the same aminoacyl-tRNA synthetase. A specific enzyme then formylates the amino group of the methionine molecule that is attached to $tRNA_f$ (Figure 30.19). The activated formyl donor in this reaction is N^{10}-formyltetrahydrofolate, a folate derivative that carries activated one-carbon units (Section 24.2). Free methionine and methionyl-$tRNA_m$ are not substrates for this transformylase.

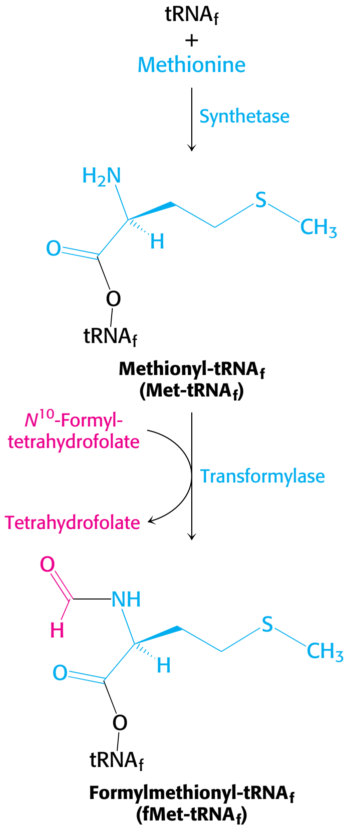

tRNAf
+
Methionine
Synthetase
H2N
S
CH3
H
O
O
tRNAf
Methionyl-tRNAf
(Met-tRNAf)
N10-Formyl-tetrahydrofolate
Transformylase
Tetrahydrofolate
O
NH
H
S
CH3
H
O
O
tRNAf
Formylmethionyl-tRNAf
(fMet-tRNAf)

FIGURE 30.19 Formylation of methionyl-tRNA. Initiator tRNA ($tRNA_f$) is first charged with methionine, and then a formyl group is transferred to the methionyl-$tRNA_f$ from N^{10}-formyltetrahydrofolate.

Formylmethionyl-tRNA$_f$ is placed in the P site of the ribosome in the formation of the 70S initiation complex

Messenger RNA and formylmethionyl-tRNA$_f$ must be brought to the ribosome for protein synthesis to begin. How is this task accomplished? Three protein *initiation factors* (IF1, IF2, and IF3) are essential. The 30S ribosomal subunit first forms a complex with IF1 and IF3 (Figure 30.20). The binding of these factors to the 30S subunit prevents it from prematurely joining the 50S subunit to form a dead-end 70S complex, devoid of mRNA and fMet-tRNA$_f$. IF1 binds near the A site and directs the fMet-tRNA$_f$ to the P site. Initiation factor 2, a member of the G-protein family, binds GTP, and the concomitant conformational change enables IF2 to associate with fMet-tRNA$_f$. The IF2–GTP–initiator-tRNA complex binds with mRNA (correctly positioned by the interaction of the Shine–Dalgarno sequence with the 16S rRNA) and the 30S subunit to form the *30S initiation complex*. Structural changes then lead to the ejection of IF1 and IF3. IF2 stimulates the association of the 50S subunit to the complex. The GTP bound to IF2 is hydrolyzed upon arrival of the 50S subunit, releasing IF2. The result is a *70S initiation complex*. The formation of the 70S initiation complex is the rate-limiting step in protein synthesis.

When the 70S initiation complex has been formed, the ribosome is ready for the elongation phase of protein synthesis. The fMet-tRNA$_f$ molecule occupies the P site on the ribosome, positioned so that its anticodon pairs with the initiating codon on mRNA. The other two sites for tRNA molecules, the A site and the E site, are empty. This interaction establishes the *reading frame* for the translation of the entire mRNA. After the initiator codon has been located, groups of three nonoverlapping nucleotides are defined.

Elongation factors deliver aminoacyl-tRNA to the ribosome

At this point, fMet-tRNA$_f$ occupies the P site, and the A site is vacant. The particular species inserted into the empty A site depends on the mRNA codon in the A site. However, the appropriate aminoacyl-tRNA does not simply leave the synthetase and diffuse to the A site. Rather, it is delivered to the A site in association with a 43-kDa protein called *elongation factor Tu* (EF-Tu), another member of the G-protein family. EF-Tu, the most abundant bacterial protein, binds aminoacyl-tRNA only in its GTP form (Figure 30.21). The binding of EF-Tu to aminoacyl-tRNA serves two functions. First, EF-Tu protects the delicate ester linkage in aminoacyl-tRNA from hydrolysis. Second, EF-Tu contributes to the accuracy of protein synthesis because GTP hydrolysis and expulsion of the EF-Tu-GDP complex from the ribosome occurs only if the pairing between the anticodon and the codon is correct. EF-Tu interacts with the 16S RNA, which monitors the accuracy of the base pairing of position 1 and 2 of the codon (p. 898). Correct codon recognition induces structural changes in the 30S subunit that activate the GTPase activity of EF-Tu, releasing EF-Tu-GDP from the ribosome. These same structural changes also rotate the aminoacyl-tRNA in the A site so that the amino acid is brought into proximity with the aminoacyl-tRNA in the P site, a process called *accommodation.* Accommodation aligns the amino acids for peptide bond formation.

Released EF-Tu is then reset to its GTP form by a second elongation factor, *elongation factor Ts.* EF-Ts induces the dissociation of GDP. GTP binds to EF-Tu, and EF-Ts concomitantly departs. It is noteworthy that *EF-Tu does not interact with fMet-tRNA$_f$*. Hence, this initiator tRNA is not

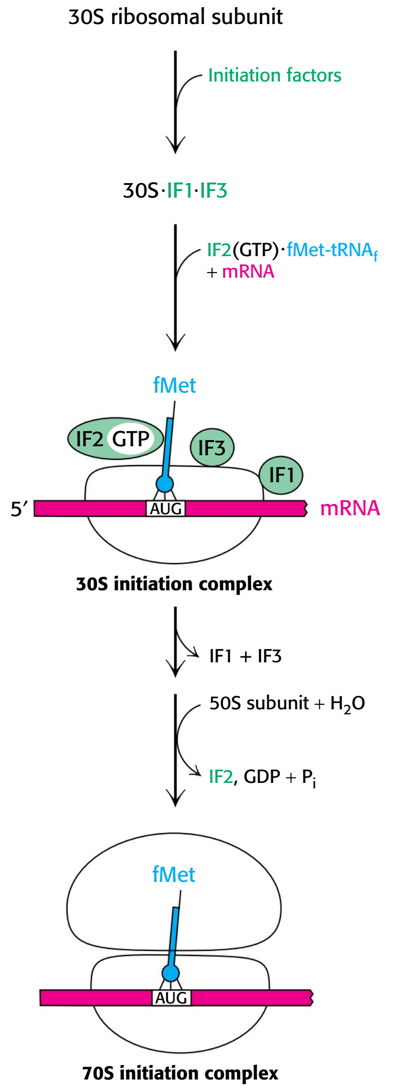


FIGURE 30.20 Translation initiation in bacteria. Initiation factors aid the assembly first of the 30S initiation complex and then of the 70S initiation complex.

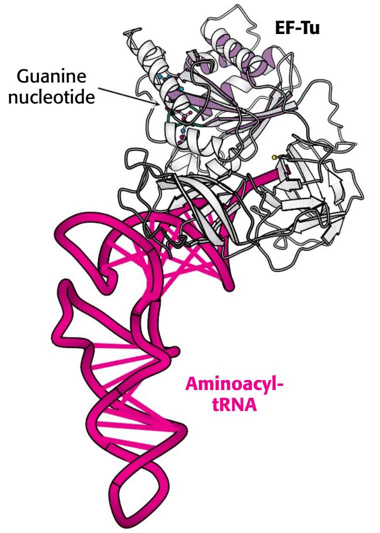


FIGURE 30.21 Structure of elongation factor Tu. The structure of a complex between elongation factor Tu (EF-Tu) and an aminoacyl-tRNA. *Notice* the P-loop NTPase domain (purple shading) at the amino-terminal end of EF-Tu. This NTPase domain is similar to those in other G proteins. [Drawn from 1B23. pdb.]

delivered to the A site. In contrast, Met-tRNA$_m$, like all other aminoacyl-tRNAs, does bind to EF-Tu. These findings account for the fact that *internal AUG codons are not read by the initiator tRNA*. Conversely, IF2 recognizes fMet-tRNA$_f$ but no other tRNA. The cycle of elongation continues until a termination codon is met.

This GTP–GDP cycle of EF-Tu is reminiscent of those of the heterotrimeric G proteins in signal transduction (Section 14.1) and the Ras proteins in growth control (Section 14.3). This similarity is due to their shared evolutionary heritage, seen in the homology of the amino-terminal domain of EF-Tu to the P-loop NTPase domains in the other G proteins. In all these related enzymes, the change in conformation between the GTP and the GDP forms leads to a change in interaction partners. A further similarity is the requirement that an additional protein catalyzes the exchange of GTP for GDP; EF-Ts catalyzes the exchange for EF-Tu, just as an activated receptor does for a heterotrimeric G protein.

Peptidyl transferase catalyzes peptide-bond synthesis

With both the P site and the A site occupied by aminoacyl-tRNA, the stage is set for the formation of a peptide bond: the formylmethionine molecule linked to the initiator tRNA will be transferred to the amino group of the amino acid in the A site. The formation of the peptide bond, one of the most important reactions in life, is a thermodynamically spontaneous reaction catalyzed by a site on the 23S rRNA of the 50S subunit called the *peptidyl transferase center*. This catalytic center is located deep in the 50S subunit near the tunnel that allows the nascent peptide to leave the ribosome.

The ribosome, which enhances the rate of peptide bond synthesis by a factor of 10^7 over the uncatalyzed reaction ($\sim 10^{-4}$ M^{-1} s^{-1}), derives much of its catalytic power from *catalysis by proximity and orientation*. The ribosome positions and orients the two substrates so that they are situated to take advantage of the inherent reactivity of an amine group (on the aminoacyl-tRNA in the A site) with an ester (on the initiator tRNA in the P site). The amino group of the aminoacyl-tRNA in the A site, in its unprotonated state, makes a nucleophilic attack on the ester linkage between the initiator tRNA and the formylmethionine molecule in the P site (Figure 30.22A). The nature

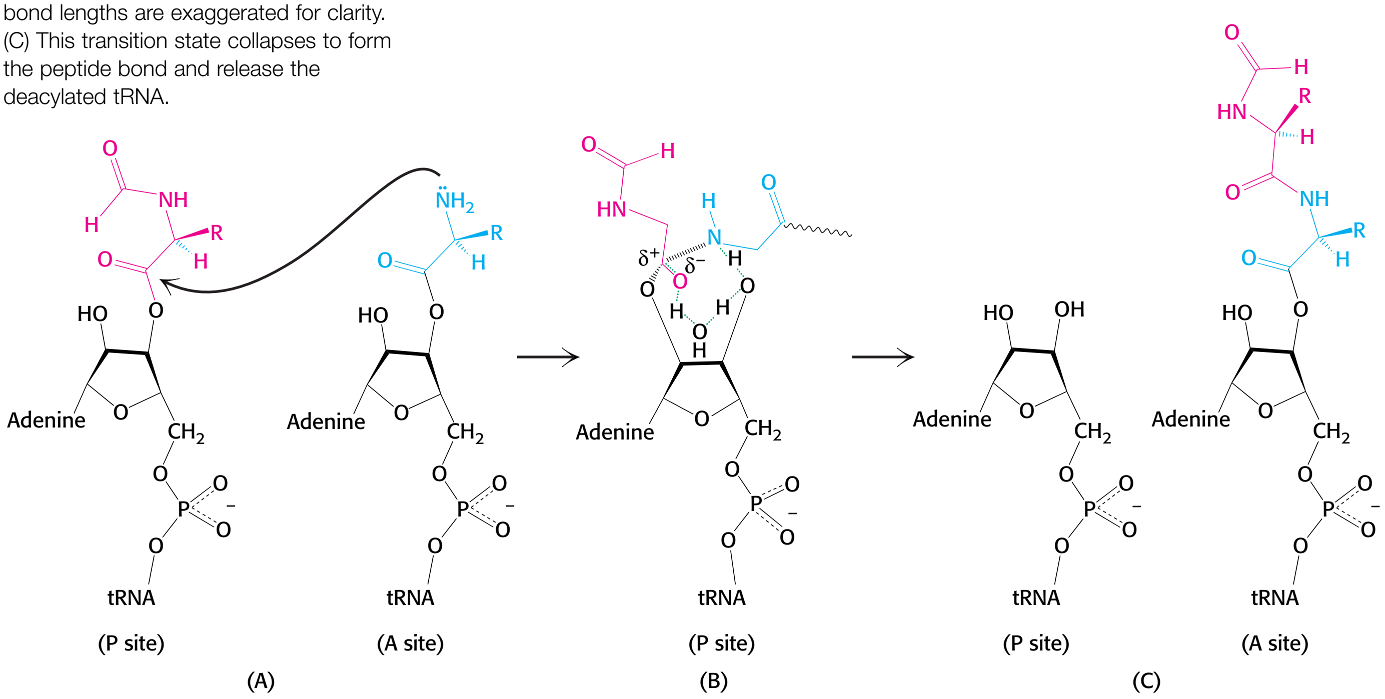


FIGURE 30.22 Peptide-bond formation. (A) The amino group of the aminoacyl-tRNA attacks the carbonyl group of the ester linkage of the peptidyl-tRNA. (B) An eight-membered transition state is formed. Note: Not all atoms are shown and some bond lengths are exaggerated for clarity. (C) This transition state collapses to form the peptide bond and release the deacylated tRNA.

of the transition state that follows the attack is not established and several models are plausible. One model proposes roles for the 2′OH of the adenosine of the tRNA in the P site and a molecule of water at the peptidyl transferase center (Figure 30.22B). The nucleophilic attack of the α-amino group generates an eight-membered transition state in which three protons are shuttled about in a concerted manner. The proton of the attacking amino group hydrogen bonds to the 2′ oxygen of ribose of the tRNA. The hydrogen of 2′OH in turn interacts with the oxygen of the water molecule at the center, which then donates a proton to the carbonyl oxygen. Collapse of the transition state with the formation of the peptide bond allows protonation of the 3′OH of the now empty tRNA in the P site (Figure 30.22C). The stage is now set for translocation and formation of the next peptide bond.

The formation of a peptide bond is followed by the GTP-driven translocation of tRNAs and mRNA

With the formation of the peptide bond, the peptide chain is now attached to the tRNA whose anticodon is in the A site on the 30S subunit. The two subunits rotate with respect to one another, and this structural change places the CCA end of the same tRNA and its peptide in the P site of the large subunit (Figure 30.23). Another aminoacyl-tRNA arrives and binds at the A site (1). Again, peptide bond synthesis occurs (2). However, protein synthesis cannot continue without the translocation of the mRNA and the tRNAs within the ribosome. *Elongation factor G* (EF-G, also called *translocase*) catalyzes the movement of mRNA, at the expense of GTP hydrolysis, by a distance of three nucleotides. Now, the next codon is positioned in the A site for interaction with the incoming aminoacyl-tRNA (3). The peptidyl-tRNA moves out of the A site into the P site on the 30S subunit and at the same time, the deacylated tRNA moves out of the P site into the E site and is subsequently released from the ribosome (4). The movement of the peptidyl-tRNA into the P site shifts the mRNA by one codon, exposing the next codon to be translated in the A site.

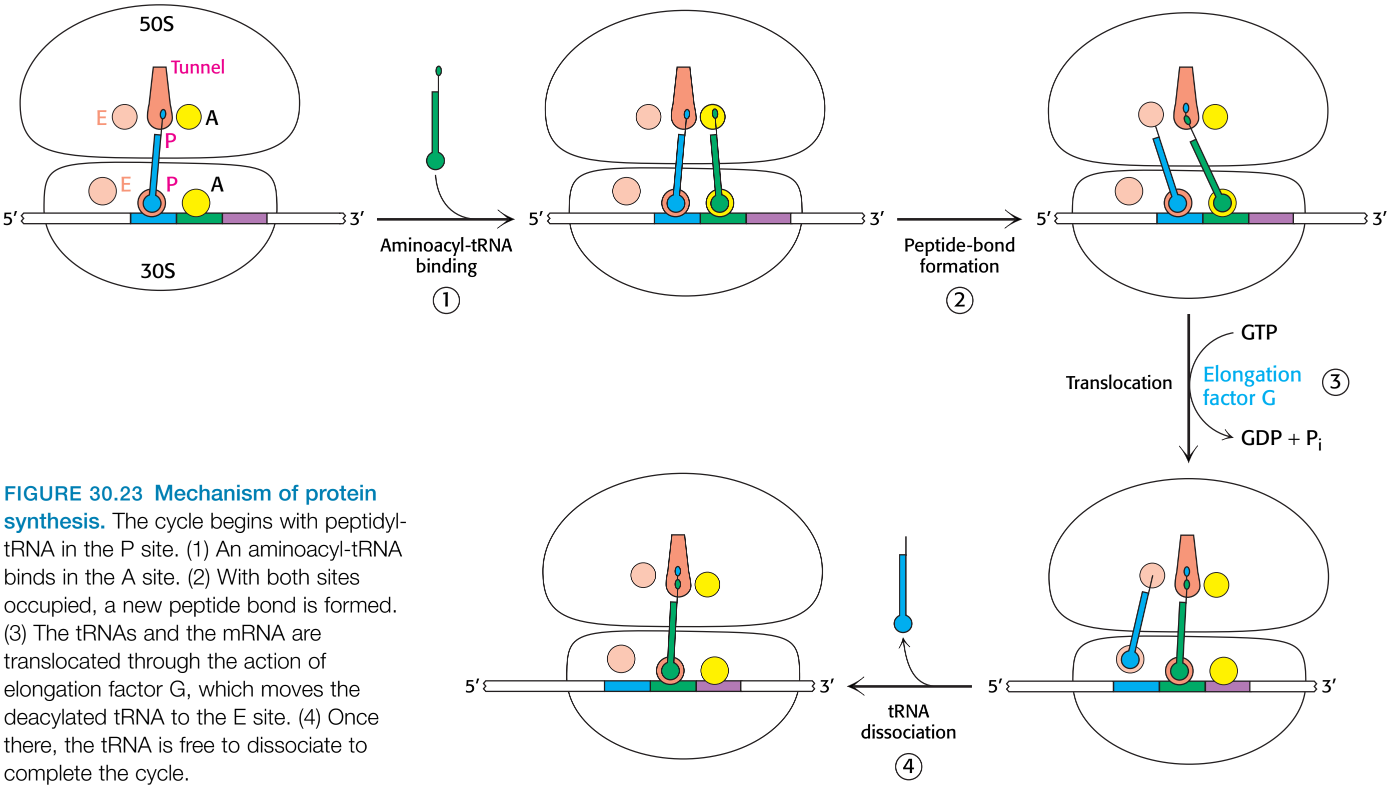


FIGURE 30.23 Mechanism of protein synthesis. The cycle begins with peptidyl-tRNA in the P site. (1) An aminoacyl-tRNA binds in the A site. (2) With both sites occupied, a new peptide bond is formed. (3) The tRNAs and the mRNA are translocated through the action of elongation factor G, which moves the deacylated tRNA to the E site. (4) Once there, the tRNA is free to dissociate to complete the cycle.

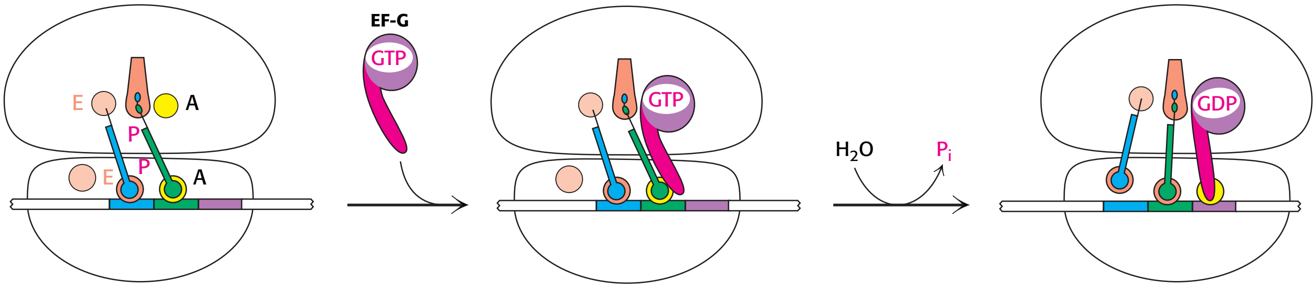


FIGURE 30.24 Translocation mechanism. In the GTP form, EF-G binds to the A site on the 50S subunit. This binding stimulates GTP hydrolysis, inducing a conformational change in EF-G that forces the tRNAs and mRNA to move through the ribosome by a distance corresponding to one codon.

The three-dimensional structure of the ribosome undergoes significant change during translocation, and evidence suggests that translocation may result from properties of the ribosome itself. However, EF-G accelerates the process. A possible mechanism for accelerating the translocation process is shown in Figure 30.24. First, EF-G in the GTP form binds to the ribosome near the A site, interacting with the 23S rRNA of the 50S subunit. The binding of EF-G to the ribosome stimulates the GTPase activity of EF-G. On GTP hydrolysis, EF-G undergoes a conformational change that displaces the peptidyl-tRNA in the A site to the P site, which carries the mRNA and the deacylated tRNA with it. The dissociation of EF-G leaves the ribosome ready to accept the next aminoacyl-tRNA into the A site.

Note that *the peptide chain remains in the P site on the 50S subunit throughout this cycle,* growing into the exit tunnel. This cycle is repeated, with mRNA translation taking place in the 5′ → 3′ direction, as new aminoacyl-tRNAs move into the A site, allowing the polypeptide to be elongated until a stop signal is found.

The direction of translation has important consequences. Recall that transcription also is in the 5′ → 3′ direction (Section 29.1). If the direction of translation were opposite that of transcription, only fully synthesized mRNA could be translated. In contrast, because the directions are the same, mRNA can be translated while it is being synthesized. In bacteria, almost no time is lost between transcription and translation. The 5′ end of mRNA interacts with ribosomes very soon after it is made, well before the 3′ end of the mRNA molecule is finished. *An important feature of bacterial gene expression is that translation and transcription are closely coupled in space and time.*

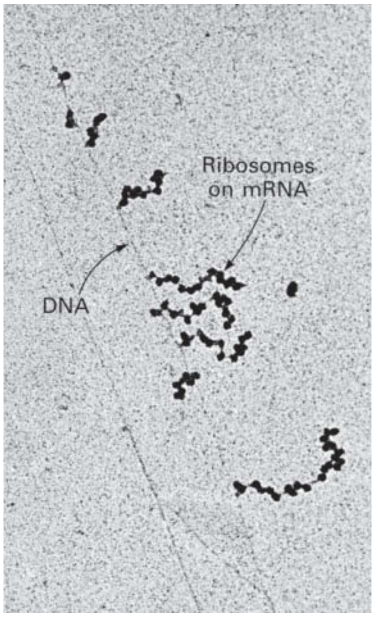


FIGURE 30.25 Polysomes. Transcription of a segment of DNA from *E. coli* generates mRNA molecules that are immediately translated by multiple ribosomes. [From O. L. Miller, Jr., B. A. Hamkalo, and C. A. Thomas, Jr. Science 169 (1970):392. Reprinted with permission from AAAS.]

Many ribosomes can be translating an mRNA molecule simultaneously. The group of ribosomes bound to an mRNA molecule is called a *polyribosome* or a *polysome* (Figure 30.25). Recent work shows that the ribosomes are arranged so as to protect the mRNA and to facilitate easy exchange of the substrates and products with the cytoplasm. The ribosomes in the polysome are in a helical array around the mRNA with the tRNA binding sites and peptide exit tunnel exposed to the cytoplasm.

Protein synthesis is terminated by release factors that read stop codons

The final phase of translation is termination. How does the synthesis of a polypeptide chain come to an end when a stop codon is encountered? No tRNAs with anticodons complementary to the stop codons—UAA, UGA, or UAG—exist in normal cells. Instead, these *stop codons are recognized by proteins called release factors* (RFs). One of these release factors, RF1, recognizes UAA or UAG. A second factor, RF2, recognizes UAA or UGA.

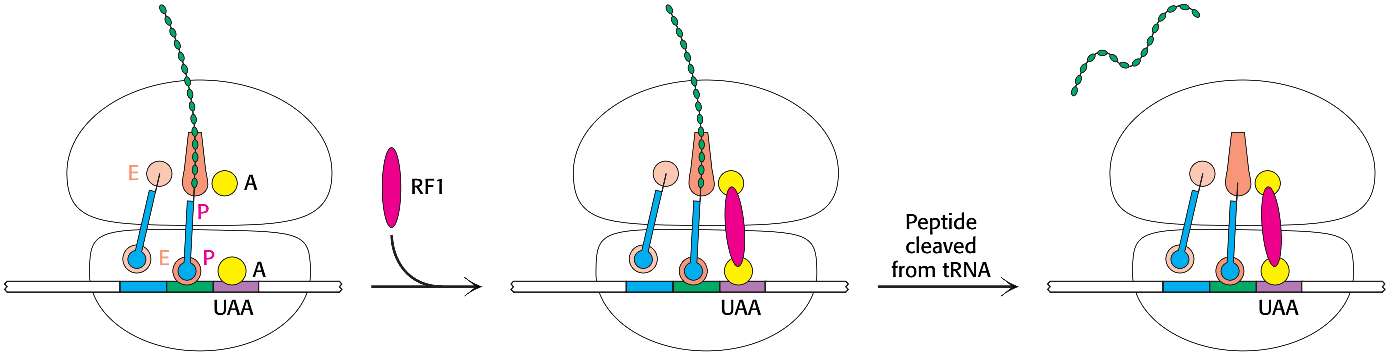
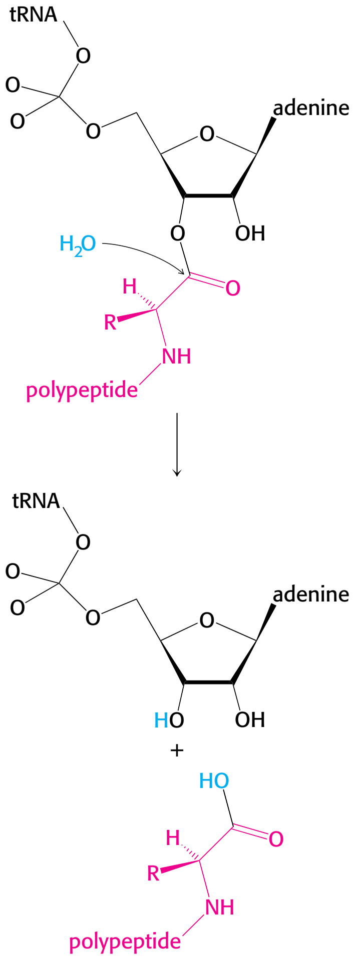


FIGURE 30.26 Termination of protein synthesis. A release factor recognizes a stop codon in the A site and stimulates the release of the completed protein from the tRNA in the P site.

A third factor, RF3, another GTPase, catalyzes the removal of RF1 or RF2 from the ribosome upon release of the newly synthesized protein.

RF1 and RF2 are compact proteins that, in eukaryotes, resemble a tRNA molecule. When bound to the ribosome, the proteins unfold to bridge the gap between the stop codon on the mRNA and the peptidyl transferase center on the 50S subunit (Figure 30.26). The RF interacts with the peptidyl transferase center using a loop containing a highly conserved glycine-glycine-glutamine (GGQ) sequence, with the glutamine methylated on the amide nitrogen atom of the R group. This modified glutamine (assisted by the peptidyl transferase) is crucial in promoting a water molecule′s attack on the ester linkage between the tRNA and the polypeptide chain, freeing the polypeptide chain. The detached polypeptide leaves the ribosome. Transfer RNA and messenger RNA remain briefly attached to the 70S ribosome until the entire complex is dissociated through the hydrolysis of GTP in response to the binding of EF-G and another factor, called the *ribosome release factor* (RRF).



30.4 Eukaryotic Protein Synthesis Differs from Bacterial Protein Synthesis Primarily in Translation Initiation

The basic plan of protein synthesis in eukaryotes and archaea is similar to that in bacteria. The major structural and mechanistic themes recur in all domains of life. However, eukaryotic protein synthesis entails more protein components than does bacterial protein synthesis, and some steps are more intricate. Some noteworthy similarities and differences are as follows:

1. *Ribosomes.* Eukaryotic ribosomes are larger. They consist of a 60S large subunit and a 40S small subunit, which come together to form an 80S particle having a mass of 4200 kDa, compared with 2700 kDa for the bacterial 70S ribosome. The 40S subunit contains an 18S RNA that is homologous to the bacterial 16S RNA. The 60S subunit contains three RNAs: the 5S RNA, which is homologous to the bacterial 5S rRNA; the 28S RNA, which is homologous to the bacterial 23S molecules; and the 5.8S RNA, which is homologous to the 5′ end of the 23S RNA of bacteria.

2. *Initiator tRNA.* In eukaryotes, the initiating amino acid is methionine rather than *N*-formylmethionine. However, as in bacteria, a special tRNA participates in initiation. This aminoacyl-tRNA is called Met-tRNA$_i$ (the subscript “i” stands for initiation).

3. *Initiation.* The initiating codon in eukaryotes is always AUG. Eukaryotes, in contrast with bacteria, do not have a specific purine-rich sequence on the 5′ side to distinguish initiator AUGs from internal ones.

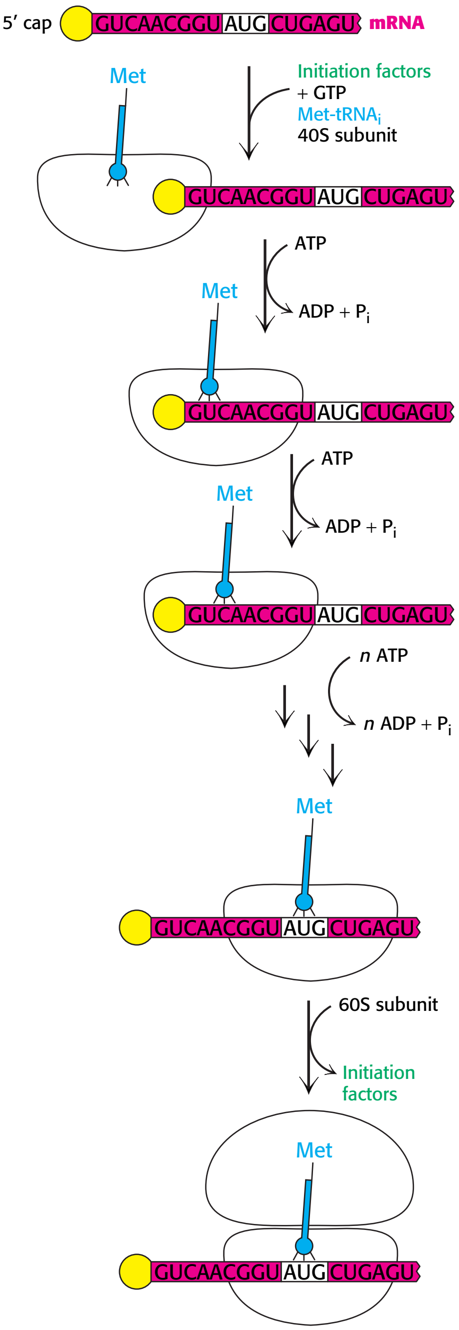


FIGURE 30.27 Eukaryotic translation initiation. In eukaryotes, translation initiation starts with the assembly of a complex on the 5′ cap that includes the 40S subunit and Met-tRNA$_i$. Driven by ATP hydrolysis, this complex scans the mRNA until the first AUG is reached. The 60S subunit is then added to form the 80S initiation complex.

Instead, the AUG nearest the 5′ end of mRNA is usually selected as the start site. Eukaryotes utilize many more initiation factors than do bacteria, and their interplay is much more complex. The prefix *eIF* denotes a eukaryotic initiation factor.

Initiation begins with the formation of a ternary complex consisting of the 40S ribosome and Met-tRNA$_i$ in association with eIF-2. The complex is called the 43S preinitiation complex (PIC). The PIC binds to the 5′ end of mRNA and begins searching for an AUG codon by moving step-by-step in the 3′ direction. Initiation factor eIF-4E binds to the 5′-cap of the mRNA (Section 29.3) and facilitates binding of PIC to the mRNA (Figure 30.27). This scanning process is catalyzed by helicases that move along the mRNA powered by ATP hydrolysis. Pairing of the anticodon of Met-tRNA$_i$ with the AUG codon of mRNA signals that the target has been found. In almost all cases, eukaryotic mRNA has only one start site and hence is the template for only a single protein. In contrast, a bacterial mRNA can have multiple Shine–Dalgarno sequences and, hence, start sites, and it can serve as a template for the synthesis of several proteins.

The difference in initiation mechanism between bacteria and eukaryotes is, in part, a consequence of the difference in RNA processing. The 5′ end of mRNA is readily available to ribosomes immediately after transcription in bacteria. In contrast, in eukaryotes pre-mRNA must be processed and transported to the cytoplasm before translation is initiated. The 5′ cap provides an easily recognizable starting point. In addition, the complexity of eukaryotic translation initiation provides another mechanism for regulation of gene expression that we shall explore further in Chapter 31.

Although most eukaryotic mRNA molecules rely on the 5′ cap to initiate protein synthesis, recent work has established that some mRNA molecules can recruit ribosomes for initiation without the use of a 5′-cap and cap-binding proteins. In these mRNAs, highly structured RNA sequences called *i*nternal *r*ibosome *e*ntry *s*ites (IRES) facilitate 40S ribosome binding to the mRNA. IRES were first discovered in the genomes of RNA viruses and have since been found in other viruses, as well as in a subset of cellular mRNA that appears to take part in development and cell stress. The molecular mechanism by which IRES function to initiate protein synthesis remains to be determined.

4. *The Structure of mRNA.* Soon after the PIC binds the mRNA, eIF-4G links eIF-4E to a protein associated with the poly(A) tail, the poly(A)-binding protein I (PABPI; Figure 30.28). Cap and tail are thus brought together to form a circle of mRNA. The benefits of circularization of mRNA remain to be determined, but a requirement for circularization may prevent translation of mRNA molecules that have lost their poly(A) tails (Section 29.3).

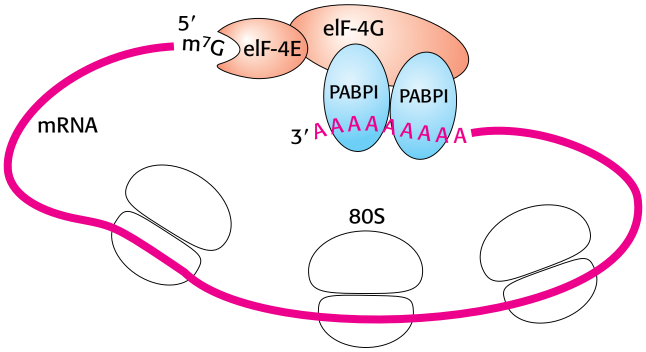


FIGURE 30.28 Protein interactions circularize eukaryotic mRNA. [Information from H. Lodish et al., *Molecular Cell Biology*, 5th ed. (W. H. Freeman and Company, 2004), Fig. 4.31.]

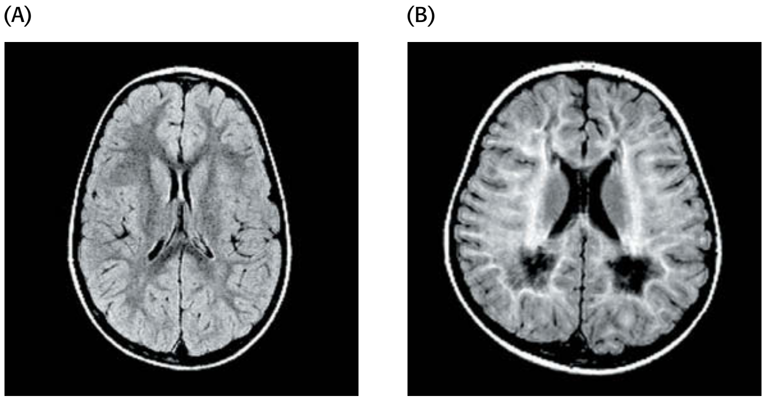


FIGURE 30.29 The effects of vanishing white matter disease. (A) In the normal brain, magnetic resonance imaging (MRI) visualizes the white matter as dark gray. (B) In the diseased brain, this MRI reveals that white matter is replaced by cerebrospinal fluid, seen as white. [(A) Macmillan Publishers Ltd: NATURE Ainsworth, Claire. "Molecular Medicine: Lost in Translation." Nature 435 (2005): 5556-558, © 2005. (B) Reprinted from Lancet Neurology, 5, M. C. van der Knaap, M. S., Pronk, J. C., Scheper, G. C., "Vanishing white matter disease" Fig. B, p. 414 © 2006, with permission from Elsevier.]

5. *Elongation and Termination.* Eukaryotic elongation factors EF1α and EF1βγ are the counterparts of bacterial EF-Tu and EF-Ts. The GTP form of EF1α delivers aminoacyl-tRNA to the A site of the ribosome, and EF1βγ catalyzes the exchange of GTP for bound GDP. Eukaryotic EF2 mediates GTP-driven translocation in much the same way as does bacterial EF-G. Termination in eukaryotes is carried out by a single release factor, eRF1, compared with two in bacteria. Release factor eIF-3 accelerates the activity of eRF-1.

6. *Organization.* The components of the translation machinery in higher eukaryotes are organized into large complexes associated with the cytoskeleton. This association is believed to facilitate the efficiency of protein synthesis. Recall that organization of elaborate biochemical processes into physical complexes is a reoccurring theme in biochemistry (Section 18.5 and Section 25.2).

Mutations in initiation factor 2 cause a curious pathological condition

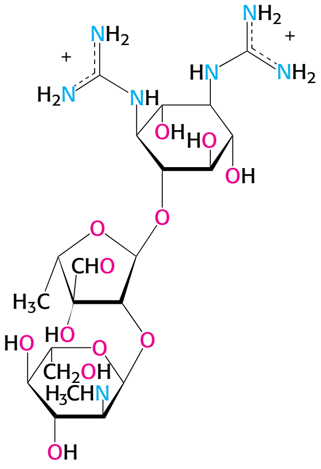
Mutations in eukaryotic initiation factor 2 result in a mysterious disease, called *vanishing white matter* (VWM) *disease,* in which nerve cells in the brain disappear and are replaced by cerebrospinal fluid (Figure 30.29). The white matter of the brain consists predominantly of nerve axons that connect the gray matter of the brain to the rest of the body. Death, resulting from fever or extended coma, can occur anywhere from a few years to decades after the onset of the disease. Symptoms usually appear in young children but can develop shortly after birth or even in adulthood. An especially puzzling aspect of the disease is its tissue specificity. A mutation in a biochemical process as fundamental to life as protein-synthesis initiation would be predicted to be lethal or at least to affect all tissues of the body. Diseases such as VWM graphically show that, although much progress has been made in biochemistry, much more research will be required to understand the complexities of health and disease.

30.5 A Variety of Antibiotics and Toxins Can Inhibit Protein Synthesis

Many chemicals that inhibit various aspects of protein synthesis have been identified. These chemicals are powerful experimental tools and clinically useful drugs.

TABLE 30.4 Antibiotic inhibitors of protein synthesis

Antibiotic	Action
Streptomycin and other aminoglycosides	Inhibit initiation and cause the misreading of mRNA (bacteria)
Tetracycline	Binds to the 30S subunit and inhibits the binding of aminoacyl-tRNAs (bacteria)
Chloramphenicol	Inhibits the peptidyl transferase activity of the 50S ribosomal subunit (bacteria)
Cycloheximide	Inhibits translocation (eukaryotes)
Erythromycin	Binds to the 50S subunit and inhibits translocation (bacteria)
Puromycin	Causes premature chain termination by acting as an analog of aminoacyl-tRNA (bacteria and eukaryotes)



Streptomycin

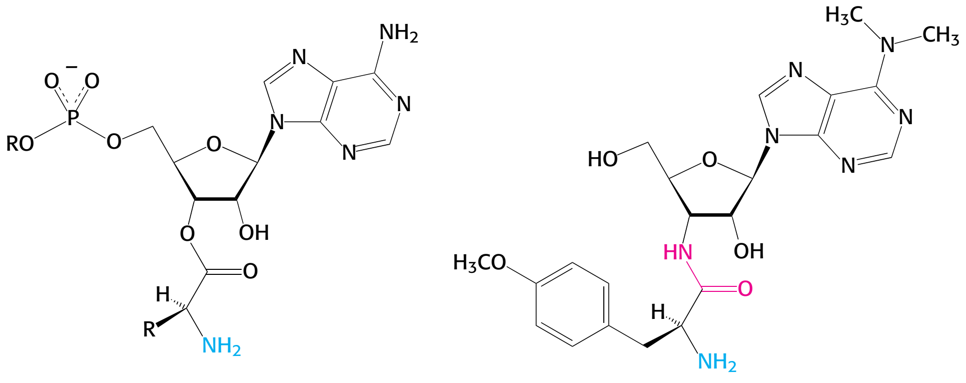
Some antibiotics inhibit protein synthesis

The differences between eukaryotic and bacterial ribosomes can be exploited for the development of antibiotics (Table 30.4). For example, the antibiotic *streptomycin,* a highly basic trisaccharide, interferes with the binding of fMet-tRNA to ribosomes in bacteria and thereby prevents the correct initiation of protein synthesis. Other *aminoglycoside antibiotics* such as neomycin, kanamycin, and gentamycin interfere with the interaction between tRNA and the 16S rRNA of the 30S subunit (p. 906) of bacterial ribosomes. *Chloramphenicol* acts by inhibiting peptidyl transferase activity. *Erythromycin* binds to the 50S subunit and blocks translocation.

The antibiotic *puromycin* inhibits protein synthesis in both bacteria and eukaryotes by causing nascent polypeptide chains to be released before their synthesis is completed. Puromycin is an analog of the terminal part of aminoacyl-tRNA (Figure 30.30). It binds to the A site on the ribosome and blocks the entry of aminoacyl-tRNA. Furthermore, puromycin contains an α-amino group. This amino group, like the one on aminoacyl-tRNA, forms a peptide bond with the carboxyl group of the growing peptide chain. The product, a peptide having a covalently attached puromycin residue at its carboxyl end, dissociates from the ribosome. No longer used medicinally, puromycin remains an experimental tool for the investigation of protein synthesis. *Cycloheximide,* another antibiotic, blocks translocation in eukaryotic ribosomes, making a useful laboratory tool for blocking protein synthesis in eukaryotic cells.

Diphtheria toxin blocks protein synthesis in eukaryotes by inhibiting translocation

Many antibiotics, harvested from bacteria for medicinal purposes, are inhibitors of bacterial protein synthesis. However, some bacteria produce protein-synthesis inhibitors that inhibit eukaryotic protein synthesis, leading to diseases such as diphtheria, which was a major cause of death in childhood before the advent of effective immunization. Symptoms include painful sore throat, hoarseness, fever, and difficulty breathing. The



Aminoacyl-tRNA

Puromycin

FIGURE 30.30 Antibiotic action of puromycin. Puromycin resembles the aminoacyl terminus of an aminoacyl-tRNA. Its amino group joins the carboxyl group of the growing polypeptide chain to form peptidyl-puromycin that dissociates from the ribosome. Peptidyl-puromycin is stable because puromycin has an amide (shown in red) rather than an ester linkage.

lethal effects of this disease are due mainly to a protein toxin produced by a phage infecting *Corynebacterium diphtheriae,* a bacterium that grows in the upper respiratory tract of an infected person. A few micrograms of diphtheria toxin is usually lethal in an unimmunized person because it inhibits protein synthesis. The toxin consists of a single polypeptide chain that is cleaved into a 21-kDa A fragment and a 40-kDa B fragment shortly after entering the cell. The role of the B-fragment in the intact protein is to bind to the cell, enabling the toxin to enter the cytoplasm of its target cell. The A fragment catalyzes the covalent modification of EF2, the elongation factor catalyzing translocation in eukaryotic protein synthesis, resulting in protein synthesis inhibition and cell death.

A single A fragment of the toxin in the cytoplasm can kill a cell. Why is it so lethal? EF2 contains *diphthamide,* an unusual amino acid residue that enhances the fidelity of codon shifting during translocation. Diphthamide is formed by a highly conserved complicated pathway that posttranslationally modifies histidine. The A fragment of the diphtheria toxin catalyzes the transfer of the ADP ribose unit of NAD^+ to the diphthamide ring (Figure 30.31). *This ADP ribosylation of a single side chain of EF2 blocks EF2's capacity to carry out the translocation of the growing polypeptide chain.* Protein synthesis ceases, accounting for the remarkable toxicity of diphtheria toxin.

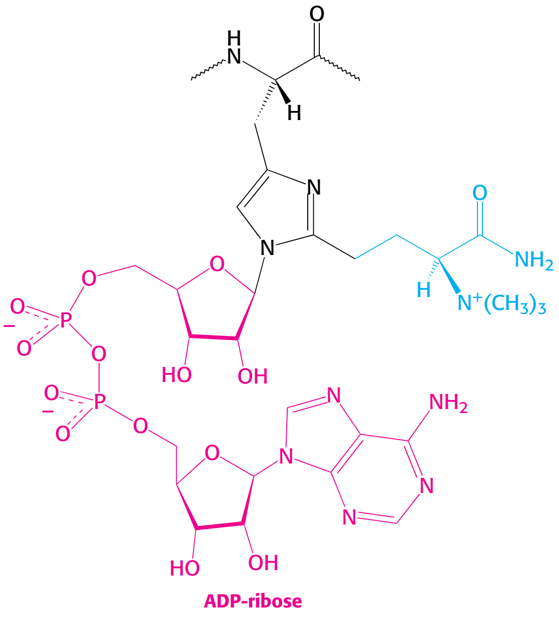


FIGURE 30.31 Blocking of translocation by diphtheria toxin. Diphtheria toxin blocks protein synthesis in eukaryotes by catalyzing the transfer of an ADP-ribose unit from NAD^+ to diphthamide, a modified amino acid residue in EF2 (translocase). Diphthamide is formed by the posttranslational modification (blue) of a histidine residue.

Ricin fatally modifies 28S ribosomal RNA

Ricin is a small protein (65 kDa) found in the seeds of the castor oil plant, *Ricinus communis* (Figure 30.32). It is indeed a deadly molecule because as little as 500 μg is lethal for an adult human being, and a single molecule can inhibit all protein synthesis in a cell, resulting in cell death.

Ricin is a heterodimeric protein composed of a catalytic A chain joined by a single disulfide bond to a B chain. The B chain allows the toxin to bind to the target cell, and this binding leads to an endocytotic uptake of the dimer and the eventual release of the A chain into the cytoplasm. The A chain, an N-glycoside hydrolyase, cleaves adenine from a particular adenosine nucleotide on the 28S rRNA that is found in all eukaryotic ribosomes. Removal of the adenine base completely inactivates the ribosome by preventing the binding of elongation factors. Thus, ricin and diphtheria toxin both act by inhibiting protein-synthesis elongation; ricin does so by covalently modifying rRNA, and diphtheria toxin does so by covalently modifying the elongation factor.

On September 7, 1978, an agent of the Soviet Union State Security Department (KGB), using a weapon built into an umbrella, embedded a small pellet containing ricin into the thigh of Bulgarian dissident Georgi Markov as he crossed Waterloo Bridge in London. Mr. Markov felt a sharp sting, as if from a bug bite. He died four days later.

FIGURE 30.32 Castor beans. The seeds of castor beans from *Ricinus communis* are a rich source of oils with a wide variety of uses, including the production of biodiesel fuels. The seeds are also rich in the toxin ricin. [Ted Kinsman/Science Source.]

30.6 Ribosomes Bound to the Endoplasmic Reticulum Manufacture Secretory and Membrane Proteins

Not all newly synthesized proteins are destined to function in the cytoplasm. Eukaryotic cells can direct proteins to internal sites such as mitochondria, the nucleus and the endoplasmic reticulum, a process called *protein targeting* or *protein sorting*. How is sorting accomplished? There are two general mechanisms by which sorting takes place. In one mechanism, the protein is synthesized in the cytoplasm, and then the completed protein is delivered to its intracellular location posttranslationally. Proteins destined for the nucleus, chloroplasts, mitochondria, and peroxisomes are delivered by this general process. The other mechanism, termed the *secretory pathway,* directs proteins into the *endoplasmic reticulum* (ER), the extensive membrane system

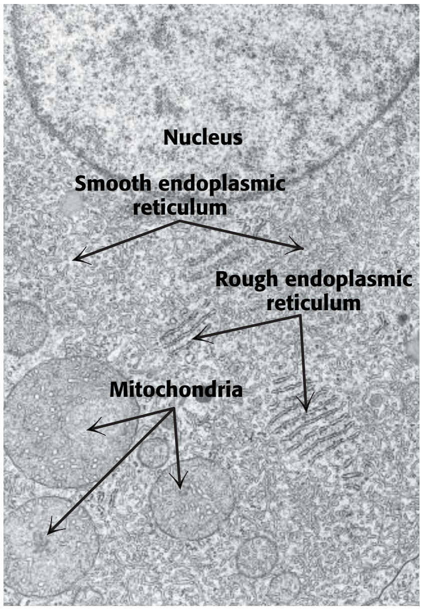


FIGURE 30.33 Ribosomes are bound to the endoplasmic reticulum. In this electron micrograph, ribosomes appear as small black dots binding to the cytoplasmic side of the endoplasmic reticulum to give a rough appearance. In contrast, the smooth endoplasmic reticulum is devoid of ribosomes. [Don W. Fawcett/Science Source]

that comprises about half the total membrane of a cell, cotranslationally—that is, while the protein is being synthesized. Approximately 30% of all proteins are sorted by the secretory pathway, including secreted proteins, residents of the ER, the Golgi complex, lysosomes, and integral membrane proteins of these organelles as well as integral plasma-membrane proteins. We will focus our attention on the secretory pathway only.

In eukaryotic cells, a ribosome remains free in the cytoplasm unless it is directed to the ER. The region that binds ribosomes is called the *rough ER* because of its studded appearance in contrast with the *smooth ER,* which is devoid of ribosomes (Figure 30.33).

Protein Synthesis Begins on Ribosomes That Are Free in the Cytoplasm

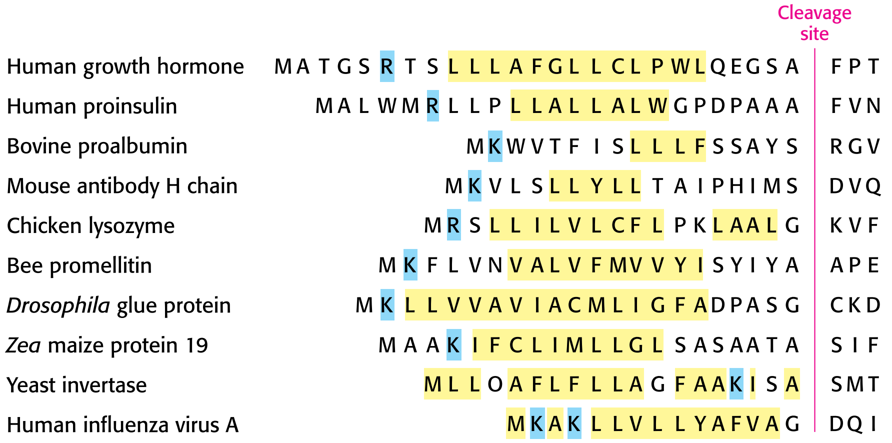
The synthesis of proteins sorted by the secretory pathway begins on free ribosomes that soon associate with the ER. Free ribosomes that are synthesizing proteins for use in the cell are identical with those attached to the ER. What is the process that directs the ribosome synthesizing a protein destined to enter the ER to bind to the ER?

Signal sequences mark proteins for translocation across the endoplasmic reticulum membrane

The synthesis of proteins destined to leave the cell or become embedded in the plasma membrane begins on a free ribosome but, shortly after synthesis begins, it is halted until the ribosome is directed to the cytoplasmic side of the ER. When the ribosome docks with the ER membrane, protein synthesis begins again. As the newly forming peptide chain exits the ribosome, it is transported, cotranslationally, through the membrane into the lumen of the ER. The translocation consists of four components.

1. *The Signal Sequence.* The signal sequence is *a sequence of 9 to 12 hydrophobic amino acid residues, sometimes containing positively charged amino acids* (Figure 30.34). This sequence, which adopts an α-helical structure, is usually near the amino terminus of the nascent polypeptide chain. The presence of the signal sequence identifies the nascent peptide as one that must cross the ER membrane. Some signal sequences are maintained in the mature protein, whereas others are cleaved by a *signal peptidase* on the lumenal side of the ER membrane.

2. *The Signal-Recognition Particle (SRP).* The signal-recognition particle recognizes the signal sequence and binds the sequence and the ribosome as soon as the signal sequence exits the ribosome. SRP then shepherds



Human growth hormone	M A T G S R T S L L L A F G L L C L P W L Q E G S A	F P T
Human proinsulin	M A L W M R L L P L L A L L A L W G P D P A A A	F V N
Bovine proalbumin	M K W V T F I S L L L F S S A Y S	R G V
Mouse antibody H chain	M K V L S L L Y L L T A I P H I M S	D V Q
Chicken lysozyme	M R S L L I L V L C F L P K L A A L G	K V F
Bee promellitin	M K F L V N V A L V F M V V Y I S Y I Y A	A P E
Drosophila glue protein	M K L L V V A V I A C M L I G F A D P A S G	C K D
Zea maize protein 19	M A A K I F C L I M L L G L S A S A A T A	S I F
Yeast invertase	M L L O A F L F L L A G F A A K I S A	S M T
Human influenza virus A	M K A K L L V L L Y A F V A G	D Q I

FIGURE 30.34 Amino-terminal signal sequences of some eukaryotic secretory and plasma-membrane proteins. The hydrophobic core (yellow) is preceded by basic residues (blue) and followed by a cleavage site (red) for signal peptidase.

the ribosome and its nascent polypeptide chain to the ER membrane. SRP is a ribonucleoprotein consisting of a 7S RNA and six different proteins (Figure 30.35). One protein, SRP54, is a GTPase that is crucial for SRP function. SRP samples all ribosomes until it locates one exhibiting a signal sequence. After SRP is bound to the signal sequence, interactions between the ribosome and the SRP occlude the elongation-factor-binding site, thereby halting protein synthesis.

3. *The SRP Receptor (SR).* The SRP–ribosome complex diffuses to the ER, where SRP binds the SRP receptor, an integral membrane protein consisting of two subunits, SRα and SRβ. SRα is, like SRP54, a GTPase.

4. *The Translocon.* The SRP–SR complex delivers the ribosome to the translocation machinery, called the *translocon,* a multisubunit assembly of integral and peripheral membrane proteins. The translocon is a protein-conducting channel. This channel opens when the translocon and ribosome bind to each other. Protein synthesis resumes with the growing polypeptide chain passing through the translocon channel into the lumen of the ER.

The interactions of the components of the translocation machinery are shown in Figure 30.36. Both the SRP54 and the SRα subunits of SR must bind GTP to facilitate the formation of the SRP–SR complex. For the SRP–SR complex to then deliver the ribosome to the translocon, the two GTP molecules—one in SRP and the other in SR—are aligned in what is

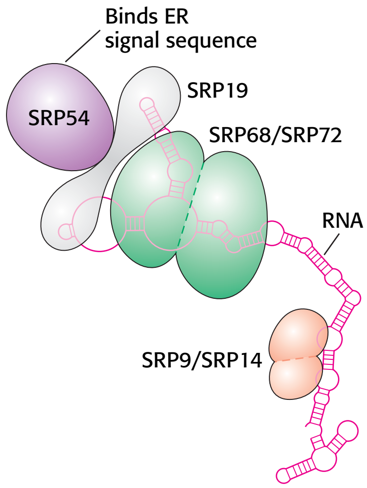
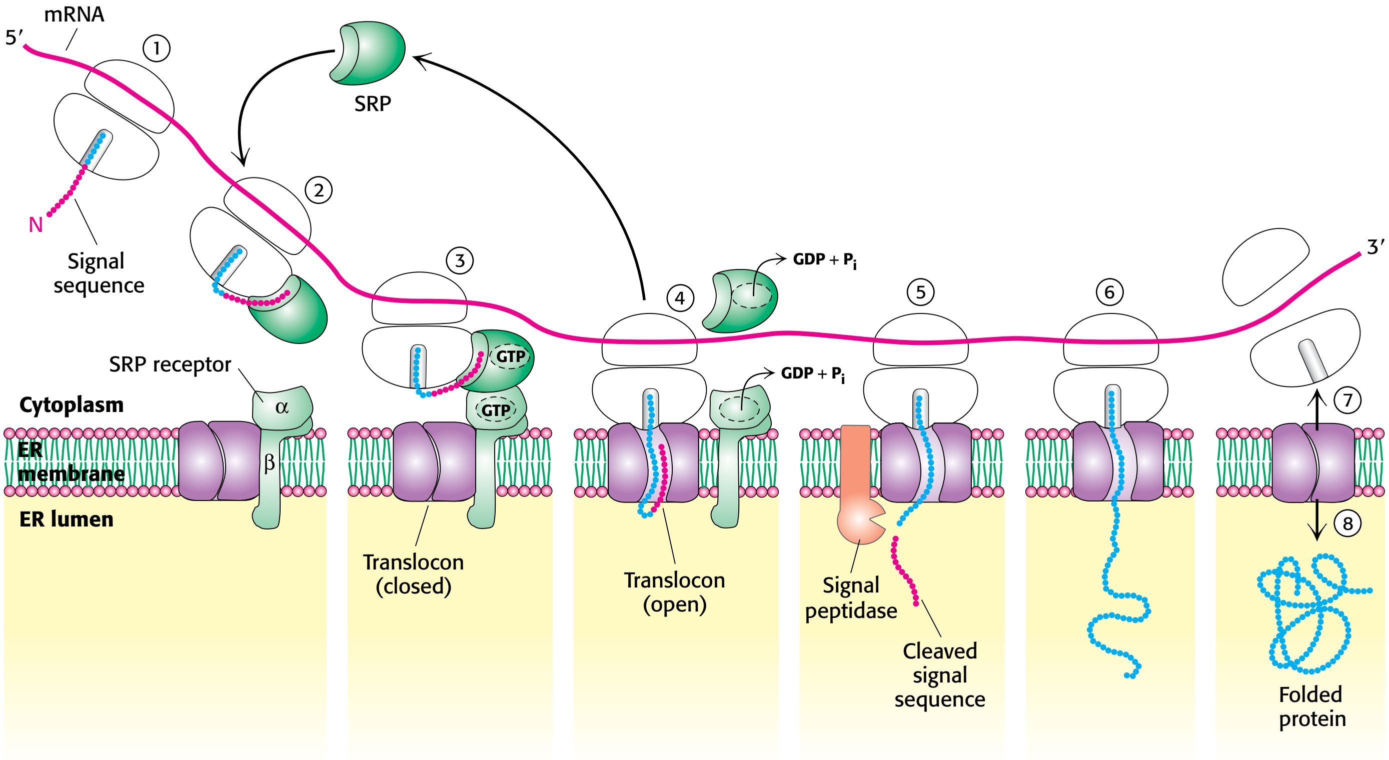


FIGURE 30.35 The signal-recognition particle. The signal-recognition particle (SRP) consists of six proteins and one 300-nucleotide RNA molecule. The RNA has a complex structure with many double-helical stretches punctuated by single-stranded regions, shown as circles. [Information from H. Lodish et al., *Molecular Cell Biology*, 5th ed. (W. H. Freeman and Company, 2004). Data from K. Strub et al., *Mol. Cell Biol*. 11:3949–3959, 1991, and S. High and B. Dobberstein, *J. Cell Biol.* 113:229–233, 1991.]

FIGURE 30.36 The SRP targeting cycle. (1) Protein synthesis begins on free ribosomes. (2) After the signal sequence has exited the ribosome, it is bound by the SRP, and protein synthesis halts. (3) The SRP–ribosome complex docks with the SRP receptor in the ER membrane. (4) The SRP and the SRP receptor simultaneously hydrolyze bound GTPs. Protein synthesis resumes and the SRP is free to bind another signal sequence. (5) The signal peptidase may remove the signal sequence as it enters the lumen of the ER. (6) Protein synthesis continues as the protein is synthesized directly into the ER. (7) On completion of protein synthesis, the ribosome is released. (8) The protein tunnel in the translocon closes. [Information from H. Lodish et al., *Molecular Cell Biology*, 5th ed. (W. H. Freeman and Company, 2004), Fig. 16.6.]



essentially an active site shared by the two proteins. The formation of the alignment is catalyzed by the 7S RNA of the SRP. After the ribosome has been passed along to the translocon, the GTPs are hydrolyzed, SRP and SR dissociate, and SRP is free to search for another signal sequence to begin the cycle anew. Thus, SRP acts catalytically. The signal peptidase, which is associated with the translocon in the lumen of the ER, removes the signal sequence from most proteins.

Transport vesicles carry cargo proteins to their final destination

As the proteins are synthesized, they fold to form their three-dimensional structures in the lumen of the ER. Some proteins are modified by the attachment of N-linked carbohydrates (Section 11.3). Finally, the proteins must be sorted and transported to their final destinations. Regardless of the destination, the principles of transport are the same. Transport is mediated by *transport vesicles* that bud off the ER (Figure 30.37). Transport vesicles from the ER carry their cargo (the proteins) to the Golgi complex, where the vesicles fuse and deposit the cargo inside the complex. There the cargo proteins are further modified—for instance, by the attachment of O-linked carbohydrates. From the Golgi complex, transport vesicles carry the cargo proteins to their final destinations, as shown in Figure 30.37.

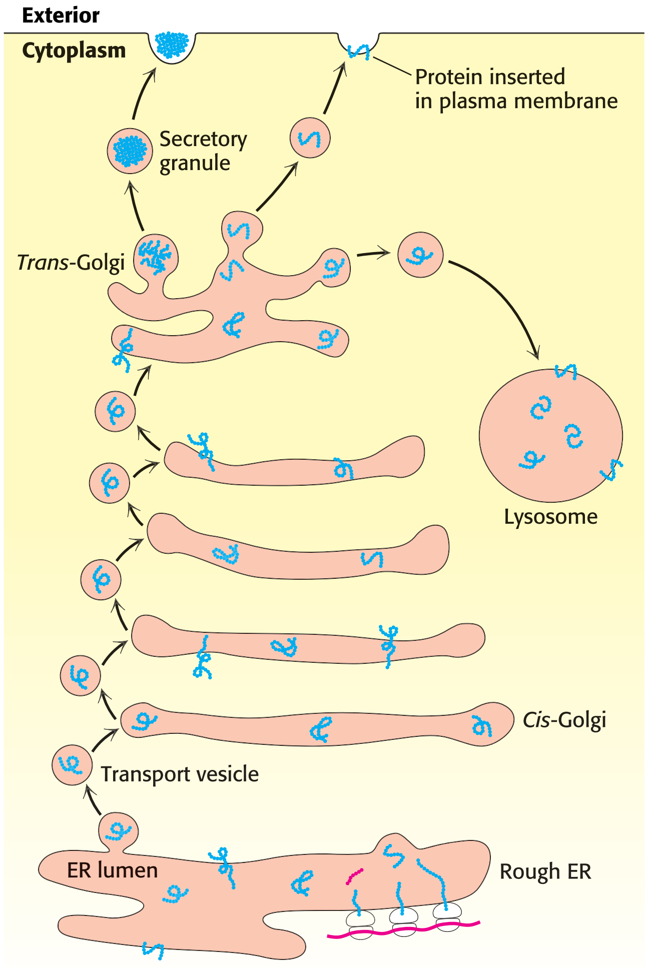


FIGURE 30.37 Protein-sorting pathways. Newly synthesized proteins in the lumen of the ER are collected into membrane buds. These buds pinch off to form transport vesicles. The transport vesicles carry the cargo proteins to the Golgi complex, where the cargo proteins are modified. Transport vesicles then carry the cargo to the final destination as directed by the v-SNARE and t-SNARE proteins.

How does a protein end up at the correct destination? A newly synthesized protein will float inside the ER lumen until it binds to an integral membrane protein called a *cargo receptor*. This binding sequesters the cargo protein into a small region of the membrane that can subsequently form a membrane bud. The bud will carry the protein to a specific destination—plasma membrane, lysosome, or cell exterior. The key to ensuring that the protein reaches the proper destination is that the protein must bind to a receptor in the ER region associated with the protein's destination. To ensure the proper match of protein with ER region, cargo receptors recognize various characteristics of the cargo protein, such as a particular amino acid sequence or an added carbohydrate.

The formation of buds is facilitated by the binding of *coat proteins* (COPs) to the cytoplasmic side of the bud. The coat proteins associate with one another to pinch off the vesicle. After the transport vesicle has formed and is released, the coat proteins are shed to reveal another integral protein called *v-SNARE* ("v" for *v*esicle). v-SNARE will bind to a particular *t-SNARE* ("t" for *t*arget) in the target membrane. This binding leads to the fusion of the transport vesicle to the target membrane, and the cargo is delivered. Thus, the assignment of identical v-SNARE proteins to the same region of the ER membrane causes an ER region to be associated with a particular destination.

SUMMARY

30.1 Protein Synthesis Requires the Translation of Nucleotide Sequences into Amino Acid Sequences

Protein synthesis is called translation because information present as a nucleic acid sequence is translated into a different language, the sequence of amino acids in a protein. This complex process is mediated by the coordinated interplay of more than a hundred macromolecules, including mRNA, rRNAs, tRNAs, aminoacyl-tRNA synthetases, and protein factors. Given that a protein typically comprises from 100 to 1000 amino acids, the frequency at which an incorrect amino acid is incorporated in the course of protein synthesis must be less than 10^{-4}. Transfer RNAs are the adaptors that make the link between a nucleic acid and an amino acid. These molecules, single chains of about 80 nucleotides, have an L-shaped structure.

30.2 Aminoacyl Transfer RNA Synthetases Read the Genetic Code

Each amino acid is activated and linked to a specific transfer RNA by an enzyme called an aminoacyl-tRNA synthetase. The synthetase links the carboxyl group of an amino acid to the 2′- or 3′-hydroxyl group of the adenosine unit of a CCA sequence at the 3′ end of the tRNA by an ester linkage. There is at least one specific aminoacyl-tRNA synthetase and at least one specific tRNA for each amino acid. A synthetase utilizes both the functional groups and the shape of its cognate amino acid to prevent the attachment of an incorrect amino acid to a tRNA. Some synthetases have a separate active site at which incorrectly linked amino acids are removed by hydrolysis. A synthetase recognizes the anticodon, the acceptor stem, and sometimes other parts of its tRNA substrate. By specifically recognizing both amino acids and tRNAs, aminoacyl-tRNA synthetases implement the instructions of the genetic code.

The codons of messenger RNA recognize the anticodons of transfer RNAs rather than the amino acids attached to the tRNAs. A codon on mRNA forms base pairs with the anticodon of the tRNA. Some tRNAs are recognized by more than one codon because pairing of the third base of a codon is less crucial than that of the other two (the wobble mechanism). There are two evolutionary distinct classes of synthetases, each recognizing 10 amino acids. The two classes recognize opposite faces of tRNA molecules.

30.3 The Ribosome Is the Site of Protein Synthesis

Protein synthesis takes place on ribosomes—ribonucleoprotein particles (about two-thirds RNA and one-third protein) consisting of large and small subunits. In *E. coli,* the 70S ribosome (2500 kDa) is made up of 30S and 50S subunits. The 30S subunit consists of 16S ribosomal RNA and 21 different proteins; the 50S subunit consists of 23S and 5S rRNA and 34 different proteins. The ribosome includes three sites for tRNA binding called the A (aminoacyl) site, the P (peptidyl) site, and the E (exit) site.

Protein synthesis takes place in three phases: initiation, elongation, and termination. In bacteria, mRNA, formylmethionyl-tRNA$_f$ (the special initiator tRNA that recognizes AUG), and a 30S ribosomal subunit come together with the assistance of initiation factors to form a 30S initiation complex. A 50S ribosomal subunit then joins this complex to form a 70S initiation complex, in which fMet-tRNA$_f$ occupies the P site of the ribosome.

Elongation factor Tu delivers the appropriate aminoacyl-tRNA to the ribosome's A site as an EF-Tu–aminoacyl-tRNA–GTP ternary complex. EF-Tu serves both to protect the aminoacyl-tRNA from premature cleavage and to increase the fidelity of protein synthesis by ensuring that the correct anticodon–codon pairing has taken place before hydrolyzing GTP and releasing aminoacyl-tRNA into the A site. A peptide bond is formed when the amino group of the aminoacyl-tRNA nucleophilically attacks the ester linkage of the peptidyl-tRNA. On peptide-bond formation, the tRNAs and mRNA are translocated for the next cycle to begin. The deacylated tRNA moves to the E site and then leaves the ribosome, and the peptidyl-tRNA moves from the A site into the P site. Elongation factor G uses the free energy of GTP hydrolysis to drive translocation. Protein synthesis is terminated by release factors, which recognize the termination codons UAA, UGA, and UAG and cause the hydrolysis of the ester bond between the polypeptide and tRNA.

30.4 Eukaryotic Protein Synthesis Differs from Bacterial Protein Synthesis Primarily in Translation Initiation

The basic plan of protein synthesis in eukaryotes is similar to that in bacteria, but there are some significant differences between them. Eukaryotic ribosomes (80S) consist of a 40S small subunit and a 60S large subunit. The initiating amino acid is again methionine, but it is not formylated. The initiation of protein synthesis is more complex in eukaryotes than in bacteria. In eukaryotes, the AUG closest to the 5′ end of mRNA is nearly always the start site. The 40S ribosome finds this site by binding to the 5′ cap and then scanning the RNA until AUG is reached. The regulation of translation in eukaryotes provides a means for regulating gene expression.

30.5 A Variety of Antibiotics and Toxins Can Inhibit Protein Synthesis

Many clinically important antibiotics function by inhibiting protein synthesis. All steps of protein synthesis are susceptible to inhibition by one antibiotic or another. Diphtheria toxin inhibits protein synthesis by covalently modifying an elongation factor, thereby preventing elongation. Ricin, a toxin from castor beans, inhibits elongation by removing a crucial adenine residue from rRNA.

30.6 Ribosomes Bound to the Endoplasmic Reticulum Manufacture Secretory and Membrane Proteins

Proteins contain signals that determine their ultimate destination. The synthesis of all proteins begins on free ribosomes in the cytoplasm. In eukaryotes, protein synthesis continues in the cytoplasm unless the nascent chain contains a signal sequence that directs the ribosome to the endoplasmic reticulum. Amino-terminal signal sequences consist of a hydrophobic segment of 9 to 12 residues preceded by a positively charged amino acid. The signal-recognition particle, a ribonucleoprotein assembly, recognizes signal sequences and brings ribosomes bearing them to the ER. A GTP–GDP cycle releases the signal sequence from SRP and then detaches SRP from its receptor. The nascent chain is then translocated across the ER membrane. Proteins are transported throughout the cell in transport vesicles.

KEY TERMS

translation (p. 893)
ribosome (p. 893)
aminoacyl-tRNA synthetase (p. 894)
codon (p. 895)
transfer RNA (tRNA) (p. 895)
anticodon (p. 895)
wobble hypothesis (p. 897)
30S subunit (p. 902)
50S subunit (p. 902)
Shine–Dalgarno sequence (p. 906)
initiation factor (p. 907)
elongation factor Tu (EF-Tu) (p. 907)
accommodation (p. 907)
elongation factor Ts (EF-Ts) (p. 907)
peptidyl transferase center (p. 908)
elongation factor G (EF-G) (translocase) (p. 909)
polysome (p. 910)
release factor (RF) (p. 910)
signal sequence (p. 916)
signal peptidase (p. 916)
signal-recognition particle (SRP) (p. 916)
SRP receptor (SR) (p. 917)
translocon (p. 917)
transport vesicle (p. 918)
coat proteins (p. 919)
v-SNARE (p. 919)
t-SNARE (p. 919)

PROBLEMS

1. *Babel fish*. Why is protein synthesis also called translation?

2. *Careful, but not too careful*. Why is it crucial that protein synthesis has an error frequency of 10^{-4}?

3. *Commonalities*. What features are common to all tRNA molecules?

4. *The ol' two step*. What two reaction steps are required for the formation of an aminoacyl-tRNA?

5. *The same but different*. Why must tRNA molecules have both unique structural features and common structural features?

6. *Charge it*. In the context of protein synthesis, what is meant by an activated amino acid?

7. *Synthetase mechanism*. The formation of isoleucyl-tRNA proceeds through the reversible formation of an enzyme-bound Ile-AMP intermediate. Predict whether ^{32}P-labeled ATP is formed from $^{32}PP_i$ when each of the following sets of components is incubated with the specific activating enzyme:

(a) ATP and $^{32}PP_i$

(b) tRNA, ATP, and $^{32}PP_i$

(c) Isoleucine, ATP, and $^{32}PP_i$

8. *1 = 2, for sufficiently large values of 1*. The energetic equivalent of two molecules of ATP is used to activate an amino acid, yet only one molecule of ATP is used. Explain.

9. *Sieves*. Using threonyl-tRNA synthetase as an example, account for the specificity of threonyl-tRNA formation.

10. *Use all available information*. Suggest a reason why there are two classes of aminoacyl-tRNA synthetases, with each class recognizing a different face of the tRNA.

11. *Going wobbly*. Explain how it is possible that some tRNA molecules recognize more than one codon.

12. *Light and heavy ribosomes*. Ribosomes were isolated from bacteria grown in a "heavy" medium (^{13}C and ^{15}N) and from bacteria grown in a "light" medium (^{12}C and ^{14}N). These 60S ribosomes were added to an in vitro system engaged in protein synthesis. An aliquot removed several hours later was analyzed by density-gradient centrifugation. How many bands of 70S ribosomes would you expect to see in the density gradient?

13. *The price of protein synthesis*. What is the smallest number of molecules of ATP and GTP consumed in the synthesis of a 200-residue protein, starting from amino acids? Assume that the hydrolysis of PP_i is equivalent to the hydrolysis of ATP for this calculation.

14. *Correct phasing*. What is meant by the phrase *reading frame?*

15. *Suppressing frameshifts*. The insertion of a base in a coding sequence leads to a shift in the reading frame, which in

most cases produces a nonfunctional protein. Propose a mutation in a tRNA that might suppress frameshifting.

16. *Tagging a ribosomal site.* Design an affinity-labeling reagent for one of the tRNA-binding sites in *E. coli* ribosomes.

17. *Viral mutation.* An mRNA transcript of a T7 phage gene contains the base sequence

↓

5′–AACUGCACGAGGUAACACAAGAUGGCU–3′

Predict the effect of a mutation that changes the red G to A.

18. *A new translation.* A transfer RNA with a UGU anticodon is enzymatically conjugated to ^{14}C-labeled cysteine. The cysteine unit is then chemically modified to alanine. The altered aminoacyl-tRNA is added to a protein-synthesizing system containing normal components except for this tRNA. The mRNA added to this mixture contains the following sequence:

5′–UUUUGCCAUGUUUGUGCU–3′

What is the sequence of the corresponding radiolabeled peptide?

19. *Two synthetic modes.* Compare and contrast protein synthesis by ribosomes with protein synthesis by the solid-phase method (Section 3.4).

20. *Triggered GTP hydrolysis.* Ribosomes markedly accelerate the hydrolysis of GTP bound to the complex of EF-Tu and aminoacyl-tRNA. What is the biological significance of this enhancement of GTPase activity by ribosomes?

21. *Blocking translation.* Devise an experimental strategy for switching off the expression of a specific mRNA without changing the gene encoding the protein or the gene's control elements.

22. *Directional problem.* Suppose that you have a protein-synthesis system that is synthesizing a protein designated A. Furthermore, you know that protein A has four trypsin-sensitive sites, equally spaced in the protein, that, on digestion with trypsin, yield the peptides A_1, A_2, A_3, A_4, and A_5. Peptide A_1 is the amino-terminal peptide, and A_5 is the carboxyl-terminal peptide. Finally, you know that your system requires 4 minutes to synthesize a complete protein A. At $t = 0$, you add all 20 amino acids, each carrying a ^{14}C label.

(a) At $t = 1$ minute, you isolate intact protein A from the system, cleave it with trypsin, and isolate the five peptides. Which peptide is most heavily labeled?

(b) At $t = 3$ minutes, what will be the order of the labeling of peptides from greatest to least?

(c) What does this experiment tell you about the direction of protein synthesis?

23. *Translator.* Aminoacyl-tRNA synthetases are the only component of gene expression that decodes the genetic code. Explain.

24. *A timing device.* EF-Tu, a member of the G-protein family, plays a crucial role in the elongation process of translation. Suppose that a slowly hydrolyzable analog of GTP were added to an elongating system. What would be the effect on the rate of protein synthesis?

25. *Not just RNA.* What are the roles of the protein factors required for protein synthesis?

26. *Membrane transport.* What four components are required for the translocation of proteins across the endoplasmic reticulum membrane?

27. *Push. Don't pull.* What is the energy source that powers the cotranslational movement of proteins across the endoplasmic reticulum?

28. *You have to know where to look.* Bacterial messenger RNAs usually contain many AUG codons. How does the ribosome identify the AUG specifying initiation?

29. *Fundamentally the same, yet …* List the differences between bacterial and eukaryotic protein synthesis.

30. *Like a border collie.* What is the role of the signal-recognition particle in protein translocation?

31. *An assembly line.* Why is the fact that protein synthesis takes place on polysomes advantageous?

32. *Match 'em*

(a) Initiation ______
(b) Elongation ______
(c) Termination ______

1. GTP
2. AUG
3. fMet
4. RRF
5. IF2
6. Shine–Dalgarno
7. EF-Tu
8. Peptidyl transferase
9. UGA
10. Transformylase

33. *Wasted effort?* Transfer RNA molecules are quite large, given that the anticodon consists of only three nucleotides. What is the purpose of the rest of the tRNA molecule?

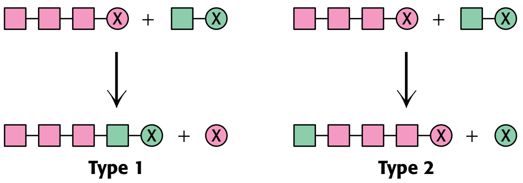
Mechanism Problems

34. *Evolutionary amino acid choice.* Ornithine is structurally similar to lysine except ornithine's side chain is one methylene group shorter than that of lysine. Attempts

to chemically synthesize and isolate ornithinyl-tRNA proved unsuccessful. Propose a mechanistic explanation. (Hint: Six-membered rings are more stable than seven-membered rings.)

Chapter Integration Problems

35. *Contrasting modes of elongation.* The two basic mechanisms for the elongation of biomolecules are represented in the adjoining illustration. In type 1, the activating group (X) is released from the growing chain. In type 2, the activating group is released from the incoming unit as it is added to the growing chain. Indicate whether each of the following biosyntheses is by means of a type 1 or a type 2 mechanism:



(a) Glycogen synthesis

(b) Fatty acid synthesis

(c) $C_5 \rightarrow C_{10} \rightarrow C_{15}$ in cholesterol synthesis

(d) DNA synthesis

(e) RNA synthesis

(f) Protein synthesis

36. *Enhancing fidelity.* Compare the accuracy of DNA replication, RNA synthesis, and protein synthesis. Which mechanisms are used to ensure the fidelity of each of these processes?

37. *Déjà vu.* Which protein in G-protein cascades plays a role similar to that of elongation factor Ts?

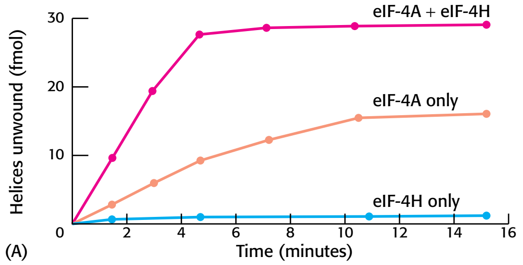
38. *Family resemblance.* Eukaryotic elongation factor 2 is inhibited by ADP ribosylation catalyzed by diphtheria toxin. What other G proteins are sensitive to this mode of inhibition?

39. *The exceptional E. coli.* In contrast with *E. coli,* most bacteria do not have a full complement of aminoacyl-tRNA synthetases. For instance, *Helicobacter pylori,* the cause of stomach ulcers, has $tRNA^{Gln}$, but no Gln-tRNA synthetase. However, glutamine is a common amino acid in *H. pylori* proteins. Suggest a means by which glutamine can be incorporated into proteins in *H. pylori.* (Hint: Glu-tRNA synthetase can misacylate $tRNA^{Gln}$.)

40. *The final step.* What aspect of primary structure allows the transfer of linear nucleic acid information into the functional three-dimensional structure of proteins?

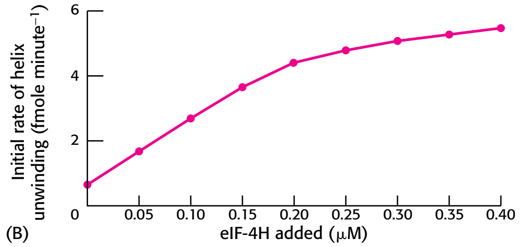
Data Interpretation Problems

41. *Helicase helper.* The initiation factor eIF-4 displays ATP-dependent RNA helicase activity. Another initiation factor, eIF-4H, has been proposed to assist the action of eIF-4. Graph A shows some of the experimental results from an assay that can measure the activity of eIF-4 helicase in the presence of eIF-4H.

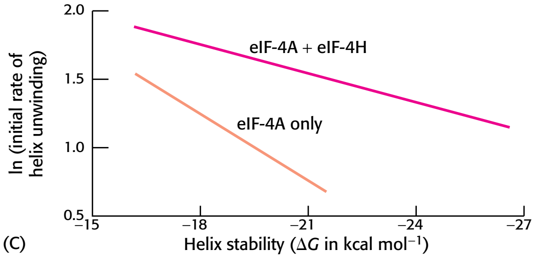


(a) What are the effects on eIF-4 helicase activity in the presence of eIF-4H?

(b) Why did measuring the helicase activity of eIF-4H alone serve as an important control?



(c) The initial rate of helicase activity of 0.2 μM of eIF-4 was then measured with varying amounts of eIF-4H (graph B). What ratio of eIF-4H to eIF-4 yielded optimal activity?

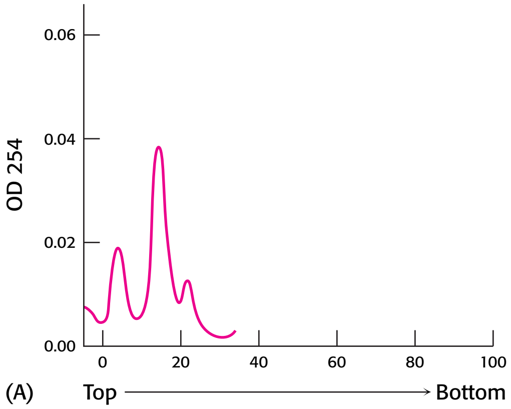


[Data from N. J. Richter, G. W. Rodgers, Jr., J. O. Hensold, and W. C. Merrick. Further biochemical and kinetic characterization of human eukaryotic initiation factor 4H., *J. Biol. Chem.* 274:35415–35424, 1999.]

(d) Next, the effect of RNA–RNA helix stability on the initial rate of unwinding in the presence and absence of eIF-4H was tested (graph C). How does the effect of eIF-4H vary with helix stability?

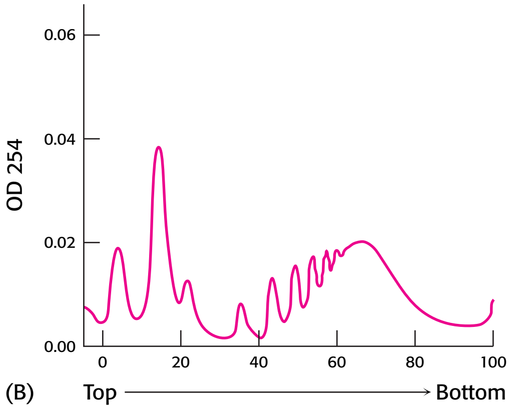
(e) How might eIF-4H affect the helicase activity of eIF-4A?

42. *Size separation*. The protein-synthesizing machinery was isolated from eukaryotic cells and briefly treated with a low concentration of RNase. The sample was then subjected to sucrose gradient centrifugation. The gradient was fractionated and the absorbance, or optical density (OD), at 254 nm was recorded for each fraction. The following plot was obtained.



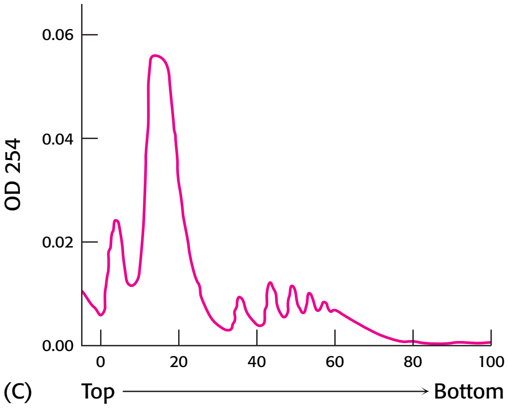
(a) What do the three peaks of absorbance in graph A represent?

The experiment was repeated except that, this time, the RNase treatment was omitted.



(b) Why is the centrifugation pattern in graph B more complex? What do the series of peaks near the bottom of the centrifuge tube represent?

Before the isolation of the protein-synthesizing machinery, the cells were grown in low concentrations of oxygen (hypoxic conditions). Again the experiment was repeated without RNase treatment (graph C).

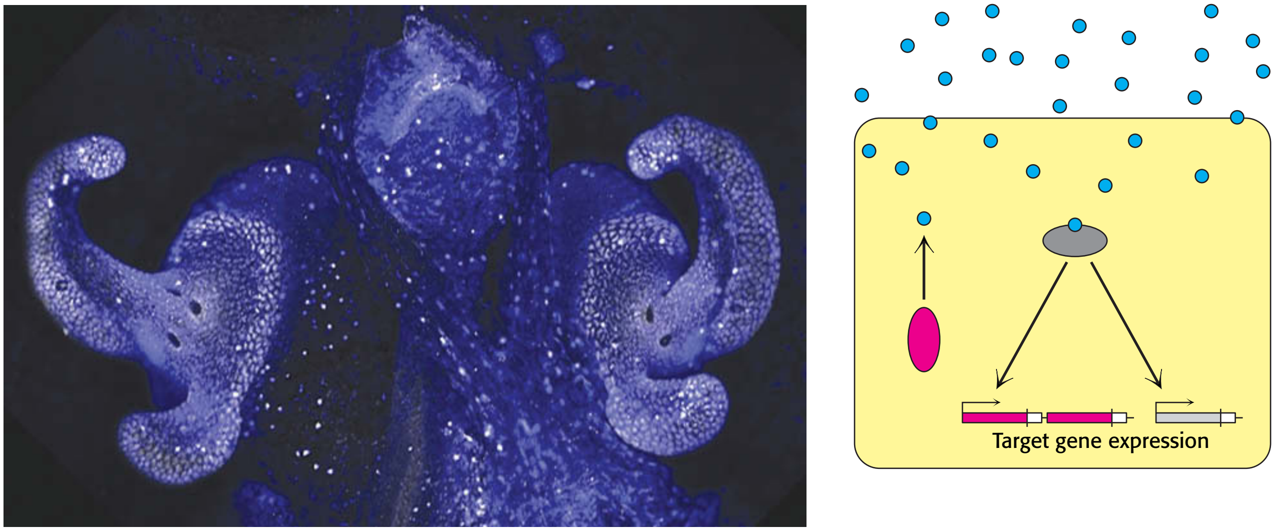


[Data from M. Koritzinsky et al., *EMBO J.* 25:1114–1125, 2006.]

(c) What is the effect of growing cells under hypoxic conditions?

CHAPTER 31

The Control of Gene Expression in Prokaryotes



Bacteria respond to changes in their environments. A micrograph of the light organ of a newly hatched squid (*Euprymna scolopes*) is shown on the left. The light spots are due to colonies of the bacteria *Vibrio fischeri* that live symbiotically within these organs. These bacteria become luminescent when they reach an appropriately high density. The density is sensed by the circuit shown on the right in which each bacterium releases a small molecule into the environment. The molecule is subsequently taken up by other bacterial cells, which start a signaling cascade that stimulates the expression of specific genes. [Spencer V. Nyholm, Eric V. Stabb, Edward G. Ruby, and Margaret J. McFall-Ngai, “Establishment of an animal–bacterial association: Recruiting symbiotic vibrios from the environment.” *PNAS* 97(2000): 10231-10235. Copyright 2000 National Academy of Sciences.]

OUTLINE

31.1 Many DNA-Binding Proteins Recognize Specific DNA Sequences

31.2 Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons

31.3 Regulatory Circuits Can Result in Switching Between Patterns of Gene Expression

31.4 Gene Expression Can Be Controlled at Posttranscriptional Levels

Even simple prokaryotic cells must respond to changes in their metabolism or in their environments. Much of this response takes place through changes in gene expression. A gene is *expressed* when it is transcribed into RNA and, for most genes, translated into proteins. Genomes comprise thousands of genes. Some of these genes are expressed all the time. These genes are subject to *constitutive expression.* Many other genes are expressed only under some circumstances—that is, under a particular set of physiological conditions. These genes are subject to *regulated expression.* For example, the level of expression of some genes in bacteria may vary more than a 1000-fold in response to the supply of nutrients or to environmental challenges.

In this chapter, we will examine gene-regulation mechanisms in prokaryotes, particularly *E. coli,* because many of these processes were first discovered in this organism. In Chapter 32, we will turn to gene-regulation mechanisms in eukaryotes. We shall see both substantial similarities and fundamental differences in comparing gene-regulatory mechanisms of the two types of organisms.

How is gene expression controlled? *Gene activity is controlled first and foremost at the level of transcription.* Whether a gene is transcribed is determined largely by the interplay between specific DNA sequences and certain proteins that bind to these sequences. Most often, these proteins repress the expression of specific genes by blocking the access of RNA polymerase to their promoters. In some cases, however, the proteins can activate the expression of specific genes. We shall learn about several different strategies that allow the coordinated regulation of sets of genes. Some genes are also controlled at stages beyond the level of transcription and we shall examine several mechanisms at these stages. Finally, we will examine several important examples of how gene expression is regulated in response to changes in the concentrations of specific molecules in the environment of prokaryotic cells.

31.1 Many DNA-Binding Proteins Recognize Specific DNA Sequences

How do regulatory systems distinguish the genes that need to be activated or repressed from genes that are constitutive? After all, the DNA sequences of genes themselves do not have any distinguishing features that would allow regulatory systems to recognize them. Instead, gene regulation depends on other sequences in the genome. In prokaryotes, these regulatory sites are close to the region of the DNA that is transcribed. Regulatory sites are usually binding sites for specific DNA-binding proteins, which can stimulate or repress gene expression. These regulatory sites were first identified in *E. coli* in studies of changes in gene expression. In the presence of the sugar lactose, the bacterium starts to express a gene encoding β-galactosidase, an enzyme that can process lactose for use as a carbon and energy source. The sequence of the regulatory site for this gene is shown in Figure 31.1. The nucleotide sequence of this site shows a nearly perfect inverted repeat, indicating that the DNA in this region has an approximate twofold axis of symmetry. Recall that cleavage sites for restriction enzymes such as EcoRV have similar symmetry properties (Section 9.3). Symmetry in such regulatory sites usually corresponds to symmetry in the protein that binds the site. *Symmetry matching is a recurring theme in protein–DNA interactions.*

5′-...TGTGTGGAATTGTGAGCGGATAACAATTTCACACA...3′
3′-...ACACACCTTAACACTCGCCTAATGTTAAAGTGTGT...5′

FIGURE 31.1 Sequence of the *lac* regulatory site. The nucleotide sequence of this regulatory site shows a nearly perfect inverted repeat, corresponding to twofold rotational symmetry in the DNA. Parts of the sequences that are related by this symmetry are shown in the same color.

To understand these protein–DNA interactions in detail, scientists examined the structure of the complex between an oligonucleotide that includes this site and the DNA-binding unit that recognizes it (Figure 31.2).

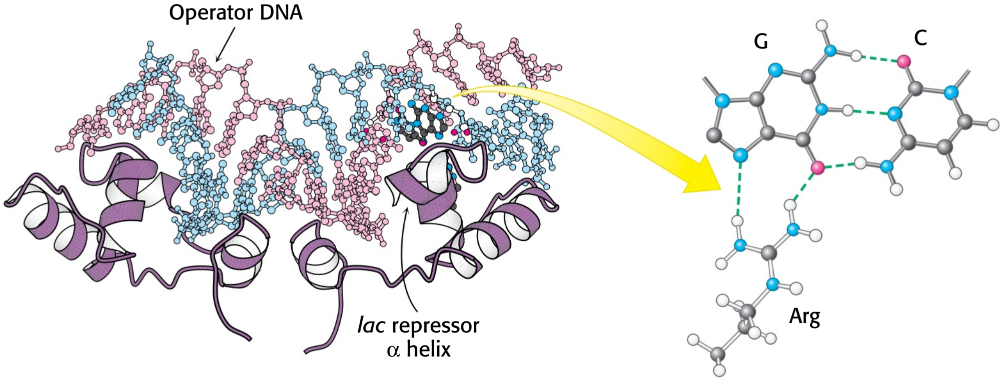


FIGURE 31.2 The *lac* repressor–DNA complex. The DNA-binding domain from a gene-regulatory protein, the *lac* repressor, binds to a DNA fragment containing its preferred binding site (referred to as operator DNA) by inserting an α helix into the major groove of operator DNA. *Notice* that a specific contact forms between an arginine residue of the repressor and a G–C base pair in the binding site. [Drawn from 1EFA.pdb.]

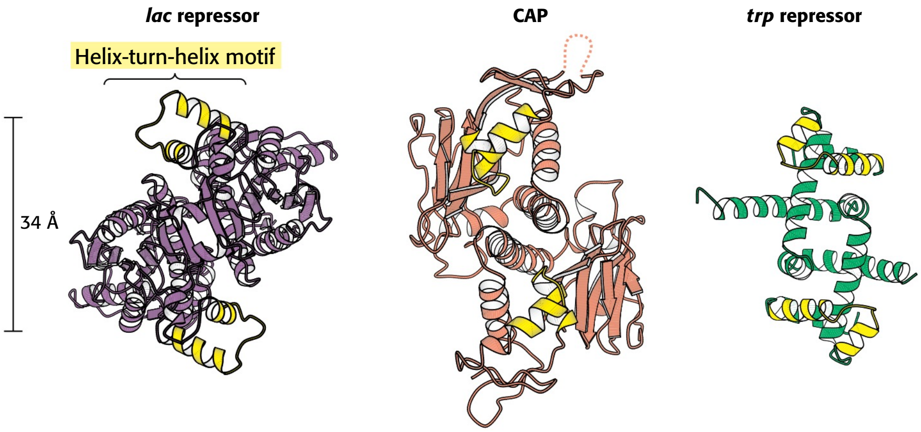


FIGURE 31.3 Helix-turn-helix motif. These structures show three sequence-specific DNA-binding proteins that interact with DNA through a helix-turn-helix motif (highlighted in yellow). *Notice* that, in each case, the helix-turn-helix units within a protein dimer are approximately 34 Å apart, corresponding to one full turn of DNA. [Drawn from 1EFA, 1RUN, and 1TRO.pdb.]

The DNA-binding unit comes from a protein called the *lac repressor,* which represses the expression of the lactose-processing gene. As expected, this DNA-binding unit binds as a dimer, and the twofold axis of symmetry of the dimer matches the symmetry of the DNA. An α helix from each monomer of the protein is inserted into the major groove of the DNA, where amino acid side chains make specific contacts with exposed edges of the base pairs. For example, the side chain of an arginine residue of the protein forms a pair of hydrogen bonds with a guanine residue of the DNA, which would not be possible with any other base. This interaction and similar ones allow the *lac* repressor to bind more tightly to this site than to the wide range of other sites present in the *E. coli* genome.

The helix-turn-helix motif is common to many prokaryotic DNA-binding proteins

Are similar strategies utilized by other prokaryotic DNA-binding proteins? The structures of many such proteins have now been determined, and amino acid sequences are known for many more. Strikingly, the DNA-binding surfaces of many, but not all, of these proteins consist of a pair of α helices separated by a tight turn (Figure 31.3). In complexes with DNA, the second of these two helices (often called the *recognition helix*) lies in the major groove, where amino acid side chains make contact with the edges of base pairs. In contrast, residues of the first helix participate primarily in contacts with the DNA backbone. *Helix-turn-helix motifs* are present on many proteins that bind DNA as dimers, and thus two of the units will be present, one on each monomer.

Although the helix-turn-helix motif is the most commonly observed DNA-binding unit in prokaryotes, not all regulatory proteins bind DNA through such units. A striking example is provided by the *E. coli* methionine repressor (Figure 31.4). This protein binds DNA through the insertion of a pair of β strands into the major groove.

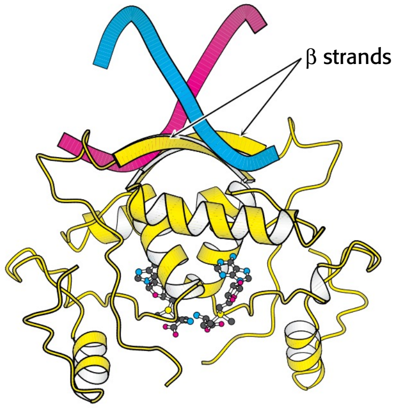


FIGURE 31.4 DNA recognition through β strands. A methionine repressor is shown bound to DNA. *Notice* that residues in β strands, rather than in α helices, participate in the crucial interactions between the protein and the DNA. [Drawn from 1CMA.pdb.]

31.2 Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons

A historically important example reveals many common principles of gene regulation by DNA-binding proteins. Bacteria such as *E. coli* usually rely on glucose as their source of carbon and energy, even when other sugars are available. However, when glucose is scarce, *E. coli* can use lactose as their carbon source, even though this disaccharide does not lie on any major metabolic pathways. An essential enzyme in the metabolism of lactose is

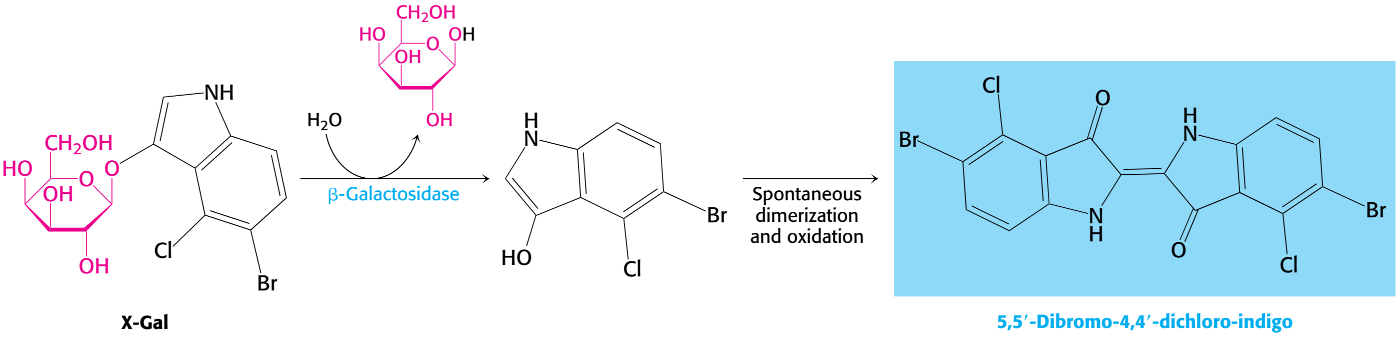
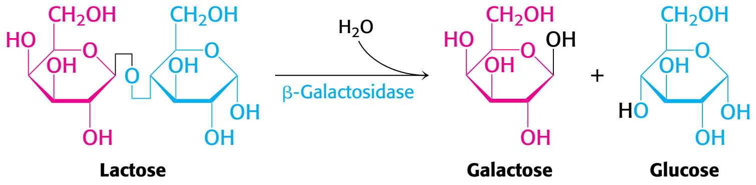

H$_2$O
β-Galactosidase
Spontaneous dimerization and oxidation
X-Gal
5,5′-Dibromo-4,4′-dichloro-indigo

FIGURE 31.5 Monitoring the β-galactosidase reaction. The galactoside substrate X-Gal produces a colored product on cleavage by β-galactosidase. The appearance of this colored product provides a convenient means for monitoring the amount of the enzyme both in vitro and in vivo.

β-galactosidase, which hydrolyzes lactose into galactose and glucose. These products are then metabolized by pathways discussed in Chapter 16.



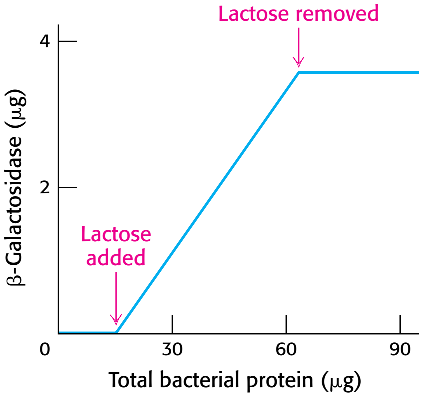


FIGURE 31.6 β-Galactosidase induction. The addition of lactose to an *E. coli* culture causes the production of β-galactosidase to increase from very low amounts to much larger amounts. The increase in the amount of enzyme parallels the increase in the number of cells in the growing culture. β-Galactosidase constitutes 6.6% of the total protein synthesized in the presence of lactose.

This reaction can be conveniently followed in the laboratory through the use of alternative galactoside substrates that form colored products such as X-Gal (Figure 31.5). An *E. coli* cell growing on a carbon source such as glucose or glycerol contains fewer than 10 molecules of β-galactosidase. In contrast, the same cell will contain several thousand molecules of the enzyme when grown on lactose (Figure 31.6). The presence of lactose in the culture medium induces a large increase in the amount of β-galactosidase by eliciting the synthesis of new enzyme molecules rather than by activating a preexisting but inactive precursor.

A crucial clue to the mechanism of gene regulation was the observation that two other proteins are synthesized in concert with β-galactosidase—namely, *galactoside permease* and *thiogalactoside transacetylase.* The permease is required for the transport of lactose across the bacterial cell membrane (Section 13.3). The transacetylase is not essential for lactose metabolism but appears to play a role in the detoxification of compounds that also may be transported by the permease. Thus, *the expression levels of a set of enzymes that all contribute to the adaptation to a given change in the environment change together.* Such a coordinated unit of gene expression is called an *operon.*

An operon consists of regulatory elements and protein-encoding genes

The parallel regulation of β-galactosidase, the permease, and the transacetylase suggested that the expression of genes encoding these enzymes is controlled by a common mechanism. François Jacob and Jacques Monod proposed the *operon model* to account for this parallel regulation as well as the results of other genetic experiments. The genetic elements of the model are a *regulator gene* that encodes a regulatory protein, a regulatory DNA sequence called *an operator site,* and a *set of structural genes* (Figure 31.7).

The regulator gene encodes a *repressor* protein that binds to the operator site. The binding of the repressor to the operator prevents transcription of the structural genes. The operator and its associated structural genes constitute

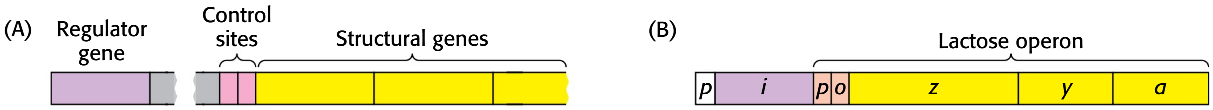


FIGURE 31.7 Operons. (A) The general structure of an operon as conceived by Jacob and Monod. (B) The structure of the lactose operon. In addition to the promoter, *p,* in the operon, a second promoter is present in front of the regulator gene, *i,* to drive the synthesis of the regulator.

the operon. For the *lactose (lac) operon,* the *i* gene encodes the repressor, *o* is the operator site, and the *z, y,* and *a* genes are the structural genes for β-galactosidase, the permease, and the transacetylase, respectively. The operon also contains a promoter site (denoted by *p*), which directs the RNA polymerase to the correct transcription initiation site. The *z, y,* and *a* genes are transcribed to give a single mRNA molecule that encodes all three proteins. An mRNA molecule encoding more than one protein is known as a *polygenic* or *polycistronic* transcript.

The *lac* repressor protein in the absence of lactose binds to the operator and blocks transcription

In the absence of lactose, the lactose operon is repressed. How does the *lac* repressor mediate this repression? The *lac repressor* exists as a tetramer of 37-kDa subunits with two pairs of subunits coming together to form the DNA-binding unit previously discussed. In the absence of lactose, the repressor binds very tightly and rapidly to the operator. When the *lac* repressor is bound to DNA, the repressor prevents RNA polymerase from transcribing the protein-coding genes inasmuch as the operator site is directly adjacent to and downstream of the promoter site where the repressor would block the progress of RNA polymerase.

How does the *lac* repressor locate the operator site in the *E. coli* chromosome? The *lac* repressor binds 4×10^6 times as strongly to operator DNA as it does to random sites in the genome. This high degree of selectivity allows the repressor to find the operator efficiently even with a large excess (4.6×10^6) of other sites within the *E. coli* genome. The dissociation constant for the repressor–operator complex is approximately 0.1 pM (10^{-13} M). The rate constant for association ($\approx 10^{10}$ M^{-1} s^{-1}) is strikingly high, indicating that the repressor finds the operator primarily by diffusing along a DNA molecule (a one-dimensional search) rather than encountering it from the aqueous medium (a three-dimensional search). This diffusion has been confirmed by studies that monitored the behavior of fluorescently labeled single molecules of *lac* repressor inside living *E. coli* cells.

Inspection of the complete *E. coli* genome sequence reveals two sites within 500 bp of the primary operator site that approximate the sequence of the operator. When one dimeric DNA-binding unit binds to the operator site, the other DNA-binding unit of the *lac* repressor tetramer can bind to one of these sites with similar sequences. The DNA between the two bound sites forms a loop. No other sites that closely match the sequence of the *lac operator* site are present in the rest of the *E. coli* genome sequence. Thus, *the DNA-binding specificity of the* lac *repressor is sufficient to specify a nearly unique site within the* E. coli *genome.*

The three-dimensional structure of the *lac* repressor has been determined in various forms. Each monomer consists of a small amino-terminal domain that binds DNA and a larger domain that mediates the formation of the dimeric DNA-binding unit and the tetramer (Figure 31.8). A pair of the amino-terminal domains come together to form the functional DNA-binding unit. Each monomer has a helix-turn-helix unit that interacts with the major groove of the bound DNA.

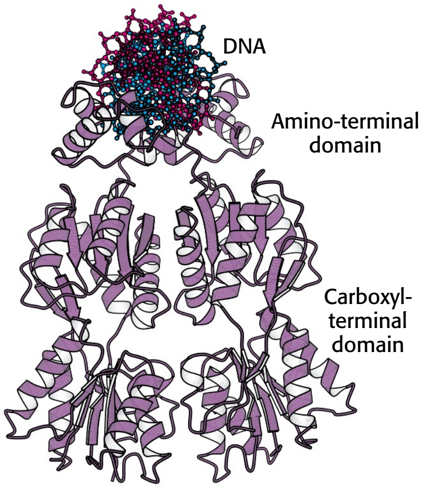
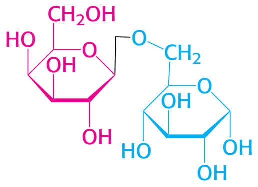
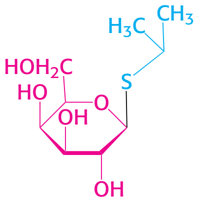


FIGURE 31.8 Structure of the *lac* repressor. A *lac* repressor dimer is shown bound to DNA. *Notice* that the amino-terminal domain binds to DNA, whereas the carboxyl-terminal domain forms a separate structure. A part of the structure that mediates the formation of *lac* repressor tetramers is not shown. [Drawn from 1EFA.pdb.]



1,6-Allolactose



Isopropylthiogalactoside (IPTG)

Ligand binding can induce structural changes in regulatory proteins

In the situation just described, glucose is present and lactose is absent, and the *lac* operon is repressed. How does the presence of lactose trigger the relief of this repression and, hence, the expression of the *lac* operon? Interestingly, lactose itself does not have this effect; rather, *allolactose,* a combination of galactose and glucose with an α-1,6 rather than an α-1,4 linkage, does. Allolactose is thus referred to as the *inducer* of the *lac* operon. Allolactose is a side product of the β-galactosidase reaction and is produced at low levels by the few molecules of β-galactosidase that are present before induction. Some other β-galactosides such as *isopropylthiogalactoside* (IPTG) are potent inducers of β-galactosidase expression, although they are not substrates of the enzyme. IPTG is useful in the laboratory as a tool for inducing gene expression in engineered bacterial strains.

The inducer triggers gene expression by preventing the *lac* repressor from binding the operator. *The inducer binds to the* lac *repressor and thereby greatly reduces the repressor's affinity for operator DNA.* An inducer molecule binds in the center of the large domain within each monomer. This binding leads to conformational changes that modify the relation between the two small DNA-binding domains (Figure 31.9). These domains can no longer easily contact DNA simultaneously, leading to a dramatic reduction in DNA-binding affinity.

Let us recapitulate the processes that regulate gene expression in the lactose operon (Figure 31.10). In the absence of inducer, the *lac* repressor is bound to DNA in a manner that blocks RNA polymerase from transcribing the *z, y,* and *a* genes. Thus, very little β-galactosidase, permease, or transacetylase are produced. The addition of lactose to the environment leads to the formation of allolactose. This inducer binds to the *lac* repressor, leading to conformational changes and the release of DNA by the *lac* repressor. With the operator site unoccupied, RNA polymerase can then transcribe the other *lac* genes and the bacterium will produce the proteins necessary for the efficient use of lactose.

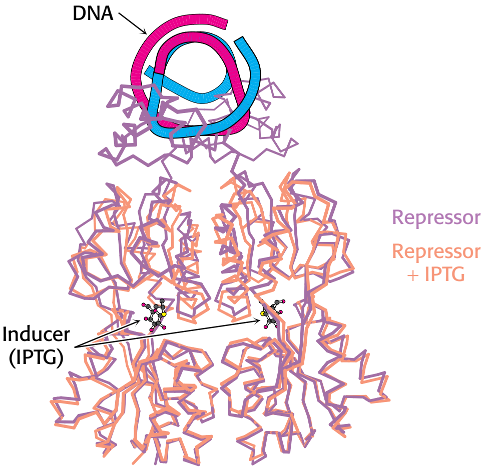


FIGURE 31.9 Effects of IPTG on *lac* repressor structure. The structure of the *lac* repressor bound to the inducer isopropylthiogalactoside (IPTG), shown in orange, is superimposed on the structure of the *lac* repressor bound to DNA, shown in purple. *Notice* that the binding of IPTG induces structural changes that alter the relation between the two DNA-binding domains so that they cannot interact effectively with DNA. The DNA-binding domains of the *lac* repressor bound to IPTG are not shown, because these regions are not well ordered in the crystals studied.

The structure of the large domain of the *lac* repressor is similar to those of a large class of proteins that are present in *E. coli* and other bacteria. This family of homologous proteins binds ligands such as sugars and amino acids at their centers. Remarkably, domains of this family are utilized by eukaryotes in taste proteins and in neurotransmitter receptors, as will be discussed in Chapter 33.

The operon is a common regulatory unit in prokaryotes

Many other gene-regulatory networks function in ways analogous to those of the *lac* operon. For example, genes taking part in purine and, to a lesser degree, pyrimidine biosynthesis are repressed by the

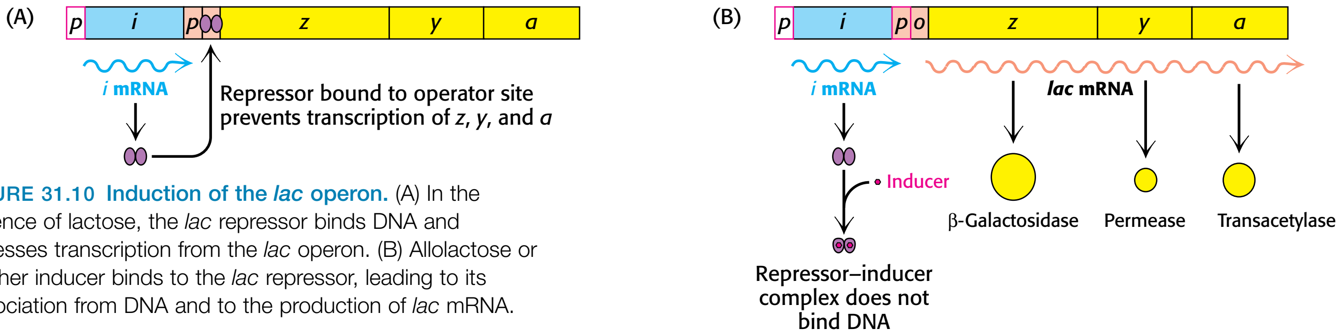

(A)
p
i
p
z
y
a
i mRNA
Repressor bound to operator site prevents transcription of z, y, and a
(B)
p
i
p
o
z
y
a
i mRNA
lac mRNA
Inducer
β-Galactosidase
Permease
Transacetylase
Repressor–inducer complex does not bind DNA

FIGURE 31.10 Induction of the *lac* operon. (A) In the absence of lactose, the *lac* repressor binds DNA and represses transcription from the *lac* operon. (B) Allolactose or another inducer binds to the *lac* repressor, leading to its dissociation from DNA and to the production of *lac* mRNA.

pur repressor. This dimeric protein is 31% identical in sequence with the *lac* repressor and has a similar three-dimensional structure. However, the behavior of the *pur* repressor is opposite that of the *lac* repressor: whereas the *lac* repressor is *released* from DNA by binding to a small molecule, *the* pur *repressor binds DNA specifically, blocking transcription, only when bound to a small molecule*. Such a small molecule is called a *corepressor*. For the *pur* repressor, the corepressor can be either guanine or hypoxanthine. The dimeric *pur* repressor binds to inverted-repeat DNA sites of the form 5′-AN**G**C**AA**N**CG**N**TT**N**C**N**T**-3′, in which the bases shown in boldface type are particularly important. Examination of the *E. coli* genome sequence reveals the presence of more than 20 such sites, regulating 19 operons and including more than 25 genes (Figure 31.11).

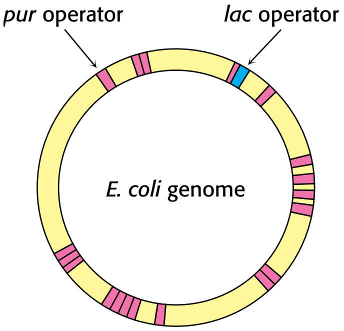


FIGURE 31.11 Binding-site distributions. The *E. coli* genome contains only a single region that closely matches the sequence of the *lac* operator (shown in blue). In contrast, 20 sites match the sequence of the *pur* operator (shown in red). Thus, the *pur* repressor regulates the expression of many more genes than does the *lac* repressor.

Because the DNA-binding sites for these regulatory proteins are short, it is likely that they evolved independently and are not related by divergence from an ancestral regulatory site. Once a ligand-regulated DNA-binding protein is present in a cell, binding sites for the protein may arise by mutation adjacent to additional genes. Binding sites for the *pur* repressor have evolved in the regulatory regions of a wide range of genes taking part in nucleotide biosynthesis. All such genes can then be regulated in a concerted manner.

The organization of prokaryotic genes into operons is useful for the analysis of completed genome sequences. Sometimes a gene of unknown function is discovered to be part of an operon containing well-characterized genes. Such associations can provide powerful clues to the biochemical and physiological functions of the uncharacterized gene.

Transcription can be stimulated by proteins that contact RNA polymerase

All the DNA-binding proteins discussed thus far function by inhibiting transcription until some environmental condition, such as the presence of lactose, is met. There are also DNA-binding proteins that stimulate transcription. One particularly well studied example is a protein in *E. coli* that stimulates the expression of catabolic enzymes.

E. coli grown on glucose, a preferred energy source, have very low levels of catabolic enzymes for metabolizing other sugars. Clearly, the synthesis of these enzymes when glucose is abundant would be wasteful. Glucose has an inhibitory effect on the genes encoding these enzymes, an effect called *catabolite repression*. It is due to the fact that glucose lowers the concentration of cyclic AMP in *E. coli*. When the cAMP concentration is high, it stimulates the concerted transcription of many catabolic enzymes by acting through a protein called the *catabolite activator protein* (CAP), which is also known as the cAMP receptor protein (CRP).

When bound to cAMP, CAP stimulates the transcription of lactose- and arabinose-catabolizing genes. CAP is a sequence-specific DNA-binding protein. Within the *lac* operon, CAP binds to an inverted repeat that is centered near position −61 relative to the start site for transcription (Figure 31.12). This site is approximately 70 base pairs from the operator site. As expected from the symmetry of the binding site, CAP functions as a dimer of identical subunits.

The CAP–cAMP complex stimulates the initiation of transcription by approximately a factor of 50. Energetically favorable contacts between CAP and RNA polymerase increase the likelihood that transcription will be initiated at sites to which the CAP–cAMP complex is bound (Figure 31.13). Thus, in regard to the *lac* operon, gene expression is maximal when the

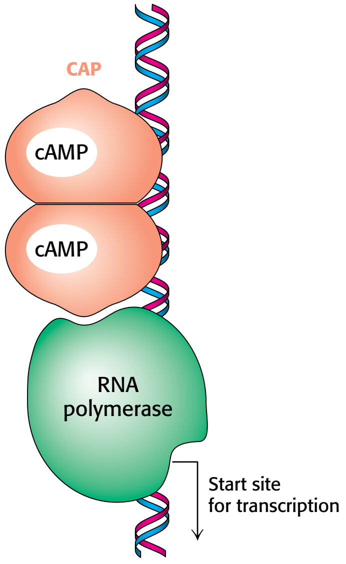


FIGURE 31.12 Binding site for catabolite activator protein (CAP). This protein binds as a dimer to an inverted repeat that is at the position −61 relative to the start site of transcription. The CAP-binding site on DNA is adjacent to the position at which RNA polymerase binds.

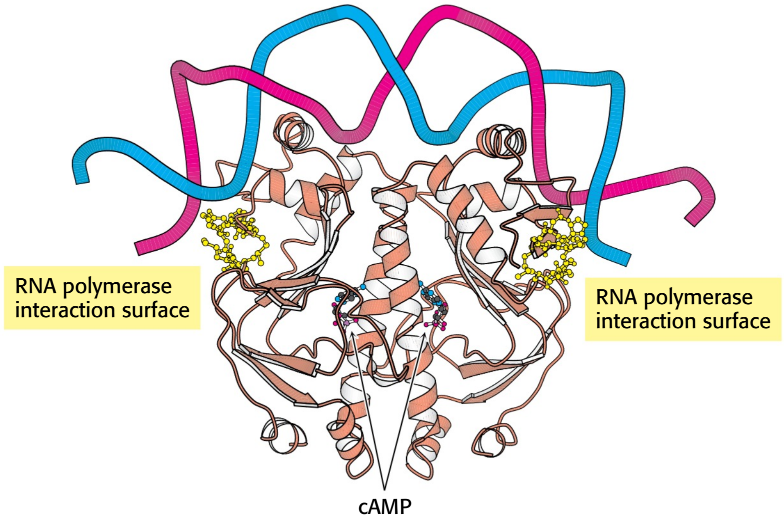

RNA polymerase interaction surface
RNA polymerase interaction surface
cAMP

FIGURE 31.13 Structure of a dimer of CAP bound to DNA. The residues shown in yellow in each CAP monomer have been implicated in direct interactions with RNA polymerase. [Drawn from 1RUN.pdb.]

binding of allolactose relieves the inhibition by the *lac* repressor and the CAP–cAMP complex stimulates the binding of RNA polymerase.

The *E. coli* genome contains many CAP-binding sites in positions appropriate for interactions with RNA polymerase. Thus, an increase in the cAMP level inside an *E. coli* bacterium results in the formation of CAP–cAMP complexes that bind to many promoters and stimulate the transcription of genes encoding a variety of catabolic enzymes.

31.3 Regulatory Circuits Can Result in Switching Between Patterns of Gene Expression

The study of viruses that infect bacteria has led to significant advances in our understanding of the processes that control gene expression. Again, sequence-specific DNA-binding proteins play key roles in these processes. Investigations of bacteriophage λ have been particularly revealing. We examined the alternative infection modes of λ phage in Chapter 5. In the lytic pathway, most of the genes in the viral genome are transcribed, initiating the production of many virus particles and leading to the eventual lysis of the bacterial cell with the concomitant release of approximately 100 virus particles. In the lysogenic pathway, the viral genome is incorporated into the bacterial DNA where most of the viral genes remain unexpressed, allowing the viral genome to be carried along as the bacteria replicate. Two key proteins and a set of regulatory sequences in the viral genome are responsible for the switch that determines which of these two pathways is followed.

The λ repressor regulates its own expression

The first protein that we shall consider is the λ *repressor*, sometimes known as the λ cI protein. This protein is key because it blocks, either directly or indirectly, the transcription of almost all genes encoded by the virus. The one exception is the gene that encodes the λ repressor itself. The λ repressor consists of an amino-terminal DNA-binding domain and a carboxyl-terminal domain that participates in protein oligomerization (Figure 31.14). This protein binds to a number of key sites in the λ phage genome. The sites of greatest interest for our present discussion are in the so-called right operator (Figure 31.15). This region includes three binding sites for the λ repressor dimer, as well as two promoters within a region of approximately 80 base pairs. One promoter drives the expression of the gene for the λ repressor itself, whereas the other drives the expression of a number of other viral genes.

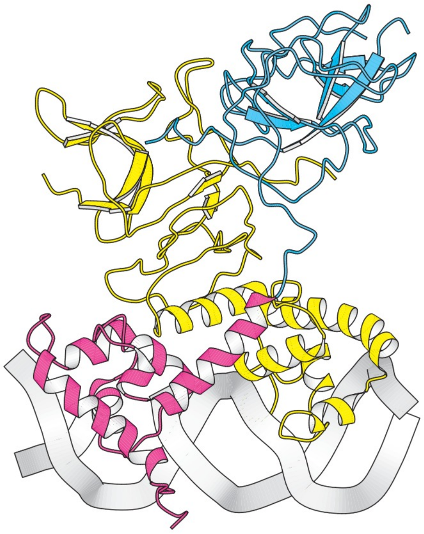


FIGURE 31.14 Structure of the λ repressor bound to DNA. The λ repressor binds to DNA as a dimer. The amino-terminal domain of one subunit is shown in red and the carboxyl-terminal domain is shown in blue. In the other subunit, both domains are shown in yellow. *Notice* how α helices on the amino-terminal domains fit into the major groove of the DNA. [Drawn from 3DBN.pdb.]

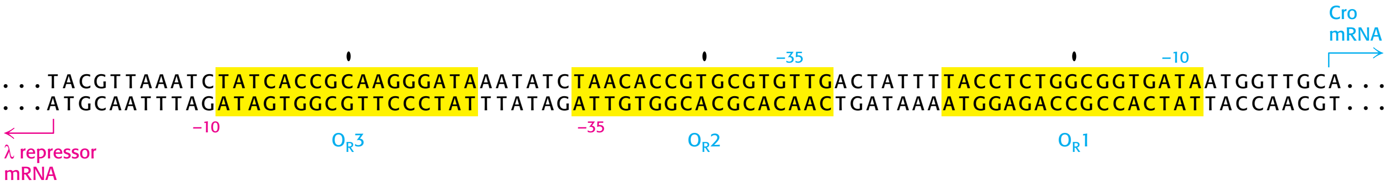

Cro
mRNA
−35
−10
. . . TACGTTAAATCTATCACCGCAAGGGATAAATATCTAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTGATAATGGTTGCA . . .
. . . ATGCAATTTAGATAGTGGCGTTCCCTATTTATAGATTGTGGCACGCACAACTGATAAAATGGAGACCGCCACTATTACCAACGT . . .
−10
−35
λ repressor
mRNA
O_R3
O_R2
O_R1

FIGURE 31.15 Sequence of the λ right operator. The three operator sites (O_R1, O_R2, and O_R3) are shaded yellow with their centers indicated. The start sites for the λ repressor mRNA and the Cro mRNA are indicated, as are their −10 and −35 sequences.

The λ repressor does not have the same affinity for the three sites; it binds the site O_R1 with the highest affinity. In addition, the binding to adjacent sites is cooperative so that, after a λ repressor dimer has bound at O_R1, the likelihood that a protein will bind to the adjacent site O_R2 increases by approximately 25-fold. Thus, when λ repressor is present in the cell at moderate concentrations, the most likely configuration has λ repressor bound at O_R1 and O_R2, but not at O_R3. In this configuration, the λ repressor dimer bound at O_R1 blocks access to the promoter on the right side of the operator sites, repressing transcription of the adjacent gene, which encodes a protein termed Cro (*c*ontroller of *r*epressor and *o*thers), while the repressor dimer at O_R2 can be in contact with RNA polymerase and stimulate transcription of the promoter that controls the transcription of the gene that encodes the λ repressor itself.

Thus, the λ repressor stimulates its own production. As the concentration of the λ repressor increases further, an additional repressor dimer can bind to the O_R3 site, blocking the other promoter and repressing the production of additional repressor. Thus, the right operator serves to maintain the λ repressor in a narrow, stable concentration range (Figure 31.16). The λ repressor also blocks other promoters in the λ phage genome so that the repressor is the only phage protein produced, which corresponds to the lysogenic state.

A circuit based on the λ repressor and Cro forms a genetic switch

What stimulates the switch to the lytic pathway? Changes such as DNA damage initiate the cleavage of the λ repressor at a specific bond between the DNA-binding and oligomerization domains. This process is mediated by the *E. coli* RecA protein (Section 28.5). After this cleavage has taken place, the affinity of the λ repressor for DNA is reduced. After the λ repressor is no longer bound to the O_R1 site, the Cro gene can be transcribed. *Cro* is a small protein that binds to the same sites as the λ repressor does, but with a different order of affinity for the three sites in the right operator. In particular, Cro has the highest affinity for O_R3. Cro bound in this site blocks the production of new repressor. The absence of repressor leads to the production of other phage genes, leading to the production of virus particles and the eventual lysis of the host cells. Thus, this genetic circuit acts as a switch with two stable states: (1) repressor high, Cro low, corresponding to the lysogenic state, and (2) Cro high, repressor low, corresponding to the lytic state (Figure 31.17). Regulatory circuits with different DNA-binding proteins controlling the expression of each other's genes constitute a common motif for controlling gene expression.

Many prokaryotic cells release chemical signals that regulate gene expression in other cells

Prokaryotic cells have been traditionally viewed as solitary single cells. However, it is becoming increasingly clear that, in many circumstances, prokaryotic cells live in complex communities, interacting with other cells

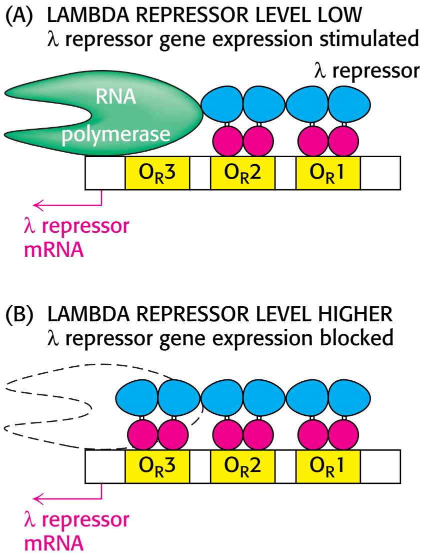

(A) LAMBDA REPRESSOR LEVEL LOW
λ repressor gene expression stimulated
λ repressor
RNA
polymerase
O_R3
O_R2
O_R1
λ repressor
mRNA
(B) LAMBDA REPRESSOR LEVEL HIGHER
λ repressor gene expression blocked
O_R3
O_R2
O_R1
λ repressor
mRNA

FIGURE 31.16 The λ repressor controls its own synthesis. (A) When λ repressor levels are relatively low, the repressor binds to sites O_R1 and O_R2 and stimulates the transcription of the gene that encodes the λ repressor itself. (B) When λ repressor levels are higher, the repressor also binds to site O_R3, blocking access to its promoter and repressing transcription from this gene.

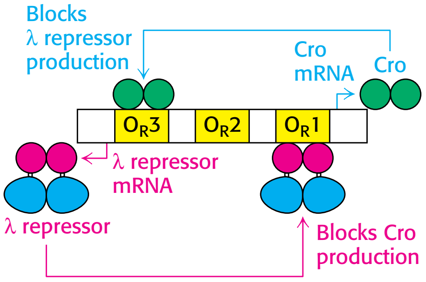
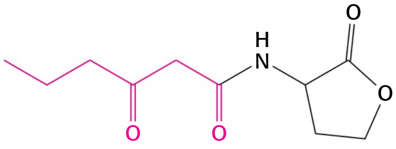

Blocks
λ repressor
production
Cro
mRNA
Cro
O_R3
O_R2
O_R1
λ repressor
mRNA
λ repressor
Blocks Cro
production

FIGURE 31.17 The λ repressor and Cro form a genetic circuit. The λ repressor blocks the production of Cro by binding most favorably to site O_R1 whereas Cro blocks the production of the λ repressor by binding most favorably to site O_R3. This circuit forms a switch that determines whether the lysogenic or the lytic pathway is followed.

of their own and different species. *These social interactions change the patterns of gene expression within the cells.*

An important type of interaction is called *quorum sensing*. This phenomenon was discovered in the bacteria *Vibrio fischeri,* a species of bacterium that can live inside a specialized light organ in the bobtail squid. In this symbiotic relation, the bacteria produce luciferase and bioluminesce, providing protection for the squid (by preventing being backlit by moonlight) in exchange for a safer place to live and reproduce. When these bacteria are grown in culture at low density, they are not bioluminescent. However, when the cell density reaches a critical level, the gene for luciferase is expressed and the cells bioluminesce. A key observation was that, when *V. fischeri* cells were transferred to a sterile medium in which other *V. fischeri* cells had been grown to high density, the cells became bioluminescent even at low cell density. This experiment revealed that a chemical, subsequently shown to be *N*-3-oxo-hexanoyl homoserine lactone (hereafter AHL for acyl-homoserine lactone), had been released into the medium where it triggered the development of the bioluminescence (Figure 31.18). This compound and other compounds that play similar roles are termed *autoinducers*.



Acyl-homoserine lactone (AHL)

FIGURE 31.18 Autoinducer structure. The structure of the acyl-homoserine lactone *N*-3-oxo-hexanoyl homoserine lactone, the autoinducer from *V. fischeri*. The autoinducers from other bacterial species can have different acyl groups (shown in red).

Cells of *V. fischeri* release the autoinducer into their environment and other *V. fischeri* cells take up the chemical. *V. fischeri* cells express a DNA-binding protein LuxR that serves as the receptor for the autoinducer. LuxR comprises two domains, one of which binds AHL and the other of which binds DNA through a helix-turn-helix motif (Figure 31.19). After the AHL concentration inside the cell has increased to an appropriate level, a substantial fraction of the LuxR molecules bind AHL. When bound to AHL, LuxR dimers bind to specific sites on DNA and increase the rate of transcription initiation at specific genes. The target genes include an operon that contains LuxA and LuxB, which together encode the luciferase enzyme, and LuxI, which produces an enzyme that catalyzes the formation of more AHL.

Because each cell produces only a small amount of the autoinducer, this regulatory system allows each *V. fischeri* cell to determine the density of the *V. fischeri* population in its environment—hence the term *quorum sensing* for this process. Studies of other prokaryotic cells are revealing an elaborate chemical language of different autoinducers (as well as autorepressors that repress specific genes). The "words" in this language include other acyl-homoserine lactone molecules with different acyl chain lengths and functionalities, as well as other distinct classes of molecules.

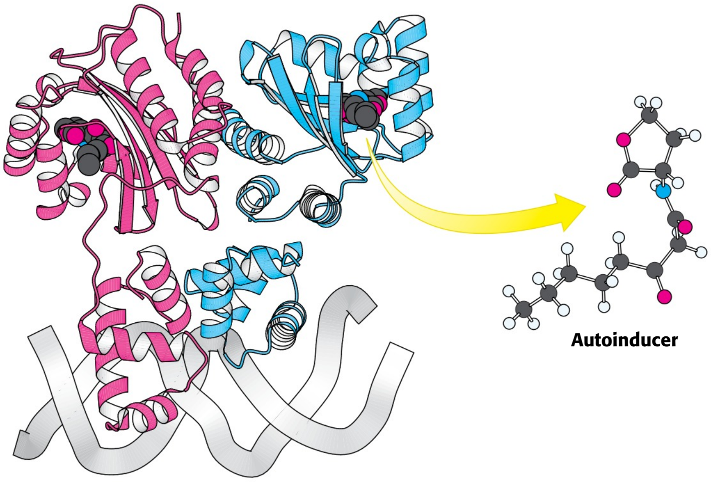


FIGURE 31.19 Quorum-sensing gene regulator. The structure of a homolog of LuxR (TraR from the bacterium *Agrobacterium tumefaciens*) is shown. *Notice* that the dimeric protein binds to DNA through an α-helical domain, whereas the autoinducer binds to a separate domain.

Biofilms are complex communities of prokaryotes

Many species of prokaryotes can be found in specialized structures termed *biofilms* that can form on surfaces. Biofilms are of considerable medical importance because organisms within them are often quite resistant to the immune response of the host as well as to antibiotics. Quorum sensing appears to play a major role in the formation of biofilms, in that cells are able to sense other cells in their environments and to promote the formation of communities with particular compositions. Some genes controlled by quorum-sensing mechanisms promote the formation of specific molecules that serve as scaffolds for the biofilm. An intriguing recent discovery is that many of the organisms that are present in biofilms on or in our bodies (perhaps 95% or more) have not been grown in culture. Through DNA-sequencing methods, we are developing a better census of our microbiome and are working toward understanding the gene-regulatory mechanisms that support these complex communities.

31.4 Gene Expression Can Be Controlled at Posttranscriptional Levels

The modulation of the rate of transcription initiation is the most common mechanism of gene regulation. However, other stages of transcription also can be targets for regulation. In addition, the process of translation provides other points of intervention for regulating the level of a protein produced in a cell. In Chapter 29, we considered riboswitches that control transcription termination (Section 29.1). Other riboswitches control gene expression by other mechanisms such as the formation of structures that inhibit translation. Additional mechanisms for posttranscriptional gene regulation have been discovered, one of which will be described here.

Attenuation is a prokaryotic mechanism for regulating transcription through the modulation of nascent RNA secondary structure

A means for regulating transcription in bacteria was discovered by Charles Yanofsky and his colleagues as a result of their studies of the tryptophan operon. This operon encodes five enzymes that convert chorismate into tryptophan. Analysis of the 5′ end of *trp* mRNA revealed the presence of a *leader sequence* of 162 nucleotides before the initiation codon of the first enzyme. The next striking observation was that bacteria produced a transcript consisting of only the first 130 nucleotides when the tryptophan level was high, but they produced a 7000-nucleotide *trp* mRNA, including the entire leader sequence, when tryptophan was scarce. Thus, when tryptophan is plentiful and the biosynthetic enzymes are not needed, transcription is abruptly broken off before any mRNA coding for the enzymes is produced. The site of termination is called the attenuator, and this mode of regulation is called *attenuation*.

Attenuation depends on features at the 5′ end of the mRNA product (Figure 31.20). The first part of the leader sequence encodes a 14-amino-acid leader peptide. Following the open reading frame for the peptide is the

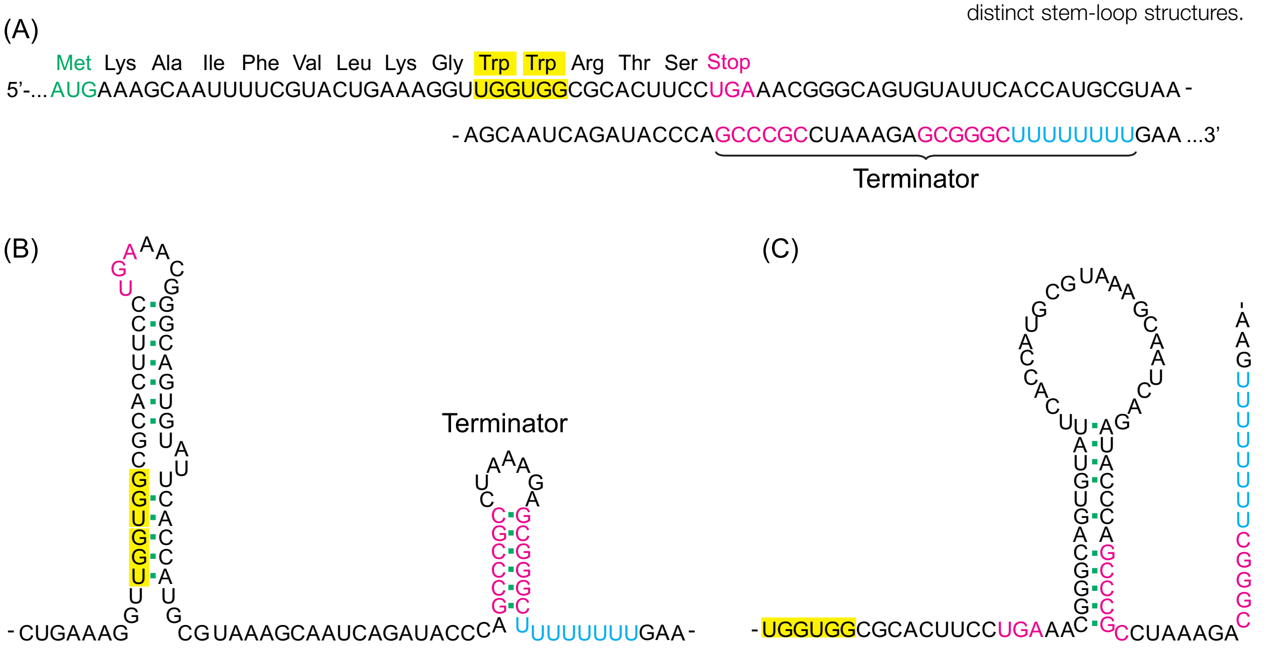

(A)
Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser Stop
5'-...AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGAAACGGGCAGUGUAUUCACCAUGCGUAA-
-AGCAAUCAGAUACCCAGCCCGCCUAAAGAGCGGGCUUUUUUUUGAA ...3'
Terminator
(B)
Terminator
-CUGAAAG
CGUAAAGCAAUCAGAUACCC
UUUUUUUGAA-
(C)
-UGGUGGCGCACUUCCUGAAA
CCUAAAGA

FIGURE 31.20 Leader region of *trp* mRNA. (A) The nucleotide sequence of the 5′ end of *trp* mRNA includes a short open reading frame that encodes a peptide comprising 14 amino acids; the leader encodes two tryptophan residues and has an untranslated attenuator region that includes a region that can form a terminator structure (red and blue). (B and C) The attenuator region can adopt either of two distinct stem-loop structures.

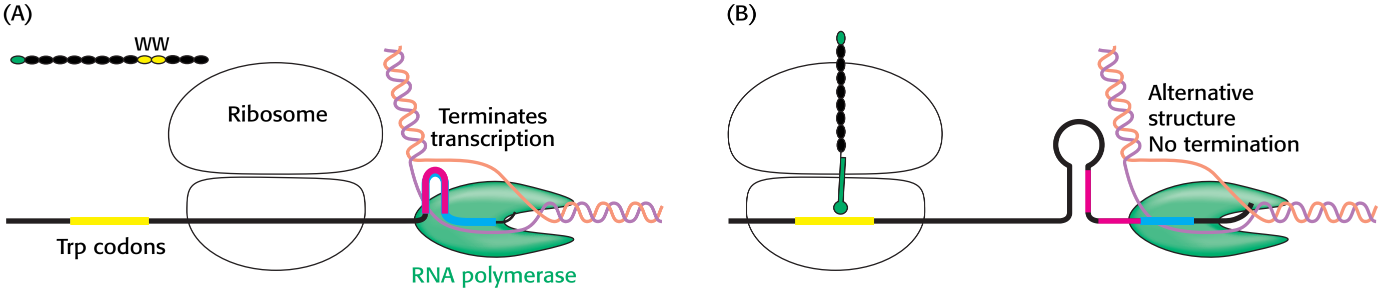


FIGURE 31.21 Attenuation. (A) In the presence of adequate concentrations of tryptophan (and, hence, Trp-tRNA), translation proceeds rapidly and an RNA structure forms that terminates transcription. (B) At low concentrations of tryptophan, translation stalls while awaiting Trp-tRNA, giving time for an alternative RNA structure to form that prevents the formation of the terminator and transcription can proceed.

attenuator, a region of RNA that is capable of forming several alternative structures. Recall that transcription and translation are tightly coupled in bacteria. Thus, the translation of the *trp* mRNA begins soon after the ribosome-binding site has been synthesized.

How does the level of tryptophan alter transcription of the *trp* operon? An important clue was the finding that the 14-amino-acid leader peptide includes two adjacent tryptophan residues. A ribosome is able to translate the leader region of the mRNA product only in the presence of adequate concentrations of tryptophan. When enough tryptophan is present, a stem-loop structure forms in the attenuator region, which leads to the release of RNA polymerase from the DNA (Figure 31.21). However, when tryptophan is scarce, transcription is terminated less frequently. Little tryptophanyl-tRNA is present, and so the ribosome stalls at the tandem UGG codons encoding tryptophan. This delay leaves the adjacent region of the mRNA exposed as transcription continues. An alternative RNA structure that does not function as a terminator is formed, and transcription continues into and through the coding regions for the enzymes. Thus, attenuation provides an elegant means of sensing the supply of tryptophan required for protein synthesis.

Several other operons for the biosynthesis of amino acids in *E. coli* also are regulated by attenuator sites. The leader peptide of each contains an abundance of the amino acid residues of the type synthesized by the operon (Figure 31.22). For example, the leader peptide for the phenylalanine operon includes 7 phenylalanine residues among 15 residues. The threonine operon encodes enzymes required for the synthesis of both threonine and isoleucine; the leader peptide contains 8 threonine and 4 isoleucine residues in a 16-residue sequence. The leader peptide for the histidine operon includes

(A)
Met - Lys - Arg - Ile - Ser - Thr - Thr - Ile - Thr - Thr - Thr - Ile - Thr - Ile - Thr - Thr -
5′ AUG AAA CGC AUU AGC ACC ACC AUU ACC ACC ACC AUC ACC AUU ACC ACA 3′

(B)
Met - Lys - His - Ile - Pro - Phe - Phe - Phe - Ala - Phe - Phe - Phe - Thr - Phe - Pro - Stop
5′ AUG AAA CAC AUA CCG UUU UUC UUC GCA UUC UUU UUU ACC UUC CCC UGA 3′

(C)
Met - Thr - Arg - Val - Gln - Phe - Lys - His - His - His - His - His - His - His - Pro - Asp -
5′ AUG ACA CGC GUU CAA UUU AAA CAC CAC CAU CAU CAC CAU CAU CCU GAC 3′

FIGURE 31.22 Leader peptide sequences. Amino acid sequences and the corresponding mRNA nucleotide sequences of the (A) threonine operon, (B) phenylalanine operon, and (C) histidine operon. In each case, an abundance of one amino acid in the leader peptide sequence leads to attenuation.

7 histidine residues in a row. In each case, low levels of the corresponding charged tRNA cause the ribosome to stall, trapping the nascent mRNA in a state that can form a structure that allows RNA polymerase to read through the attenuator site. Evolution has apparently converged on this strategy repeatedly as a mechanism for controlling amino acid biosynthesis.

The examples of prokaryotic gene regulation that we have been describing come from bacteria, as opposed to archaea. The transcriptional apparatus present in archaea shares many features with that found in eukaryotes. This commonality is frequently interpreted to suggest that eukaryotes evolved after a cell fusion event in which a bacterial cell was engulfed by an archaeal cell. Nonetheless, the key principles of gene regulation such as the occurrence of operons and the roles of DNA-binding proteins in directly blocking or stimulating RNA polymerase are the same in bacteria and archaea, perhaps because of similar genome sizes. As we shall see in the next chapter, eukaryotic cells with larger genomes and, often, many distinct cell types, utilize quite different strategies.

SUMMARY

31.1 Many DNA-Binding Proteins Recognize Specific DNA Sequences

The regulation of gene expression depends on the interplay between specific sequences within the genome and proteins that bind specifically to these sites. Specific DNA-binding proteins recognize regulatory sites that usually lie adjacent to the genes whose transcription is regulated by these proteins. The proteins of the largest family contain a helix-turn-helix motif. The first helix of this motif inserts into the major groove of DNA and makes specific hydrogen-bonding and other contacts with the edges of the base pairs.

31.2 Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons

In prokaryotes, many genes are clustered into operons, which are units of coordinated genetic expression. An operon consists of control sites (an operator and a promoter) and a set of structural genes. In addition, regulator genes encode proteins that interact with the operator and promoter sites to stimulate or inhibit transcription. The treatment of *E. coli* with lactose induces an increase in the production of β-galactosidase and two additional proteins that are encoded in the lactose operon. In the absence of lactose or a similar galactoside inducer, the *lac* repressor protein binds to an operator site on the DNA and blocks transcription. The binding of allolactose, a derivative of lactose, to the *lac* repressor induces a conformational change that leads to dissociation from DNA. RNA polymerase can then move through the operator to transcribe the *lac* operon.

Some proteins activate transcription by directly contacting RNA polymerase. For example, cyclic AMP stimulates the transcription of many catabolic operons by binding to the catabolite activator protein. The binding of the cAMP–CAP complex to a specific site in the promoter region of an inducible catabolic operon enhances the binding of RNA polymerase and the initiation of transcription.

31.3 Regulatory Circuits Can Result in Switching Between Patterns of Gene Expression

The study of bacterial viruses, particularly bacteriophage λ, has revealed key aspects of gene-regulatory networks. Bacteriophage λ can develop by either a lytic or a lysogenic pathway. A key regulatory protein, the λ repressor, regulates its own expression, promoting transcription of the gene that encodes the repressor when repressor levels are low and blocking transcription of the gene when levels are high. This behavior depends on the λ right operator, which includes three sites to which λ repressor dimers can bind. Cooperative binding of the λ repressor to two of the sites stabilizes the state in which two λ repressor dimers are bound. The Cro protein binds to the same sites as does the λ repressor, but with reversed affinities. When Cro is present at sufficient concentrations, it blocks the transcription of the gene for the λ repressor while allowing the transcription of its own gene. Thus, these two proteins and the operator form a genetic switch that can exist in either of two states.

Some prokaryotic species participate in quorum sensing. This process includes the release of chemicals called autoinducers into the medium surrounding the cells. These autoinducers are often, but not always, acyl homoserine lactones. Autoinducers are taken up by surrounding cells. When the autoinducer concentration reaches an appropriate level, it is bound by receptor proteins that activate the expression of genes, including those that promote the synthesis of more autoinducer. These chemically mediated social interactions allow these prokaryotes to change their gene-expression patterns in response to the number of other cells in their environments. Biofilms are complex communities of prokaryotes that are promoted by quorum-sensing mechanisms.

31.4 Gene Expression Can Be Controlled at Posttranscriptional Levels

Gene expression can also be regulated at the level of translation. In prokaryotes, many operons important in amino acid biosynthesis are regulated by attenuation, a process that depends on the formation of alternative structures in mRNA, one of which favors the termination of transcription. Attenuation is mediated by the translation of a leader region of mRNA. A ribosome stalled by the absence of an aminoacyl-tRNA needed to translate the leader mRNA alters the structure of mRNA, allowing RNA polymerase to transcribe the operon beyond the attenuator site.

KEY TERMS

helix-turn-helix motif (p. 927)
β-galactosidase (p. 928)
operon model (p. 928)
operator site (p. 928)
repressor (p. 928)
lac repressor (p. 929)
lac operator (p. 929)
inducer (p. 930)
isopropylthiogalactoside (IPTG) (p. 930)
pur repressor (p. 931)
corepressor (p. 931)
catabolite repression (p. 931)
catabolite activator protein (CAP) (p. 931)
λ repressor (p. 932)
CRO (p. 933)
quorum sensing (p. 934)
autoinducer (p. 934)
biofilm (p. 934)
attenuation (p. 935)

PROBLEMS

1. *Missing genes.* Predict the effects of deleting the following regions of DNA:

(a) The gene encoding *lac* repressor

(b) The *lac* operator

(c) The gene encoding CAP

2. *Minimal concentration.* Calculate the concentration of *lac* repressor, assuming that one molecule is present per cell. Assume that each *E. coli* cell has a volume of $10^{-12}\ \text{cm}^3$. Would you expect the single molecule to be free or bound to DNA?

3. *Counting sites.* Calculate the expected number of times that a given 8-base-pair DNA site should be present in the *E. coli* genome. Assume that all four bases are equally probable. Repeat for a 10-base-pair site and a 12-base-pair site.

4. *The same but not the same.* The *lac* repressor and the *pur* repressor are homologous proteins with very similar three-dimensional structures, yet they have different effects on gene expression. Describe two important ways in which the gene-regulatory properties of these proteins differ.

5. *The opposite direction.* Some compounds called anti-inducers bind to repressors such as the *lac* repressor and inhibit the action of inducers; that is, transcription is repressed and higher concentrations of inducer are required to induce transcription. Propose a mechanism of action for anti-inducers.

6. *Inverted repeats.* Suppose that a nearly perfect 20-base-pair inverted repeat is observed in a DNA sequence. Provide two possible explanations.

7. *Broken operators.* Consider a hypothetical mutation in O_R2 that blocks both λ repressor and Cro binding. How would this mutation affect the likelihood of bacteriophage λ entering the lytic phase?

8. *Promoters.* Compare the −10 and −35 sequences for the λ repressor and Cro genes in the right operator. How many differences are there between these sequences?

9. *Positive and negative feedback.* What is the effect of an increased Cro concentration on the expression of the gene for the λ repressor? Of an increased concentration of λ repressor on the expression of the Cro gene? Of an increased concentration of λ repressor on the expression of the λ repressor gene?

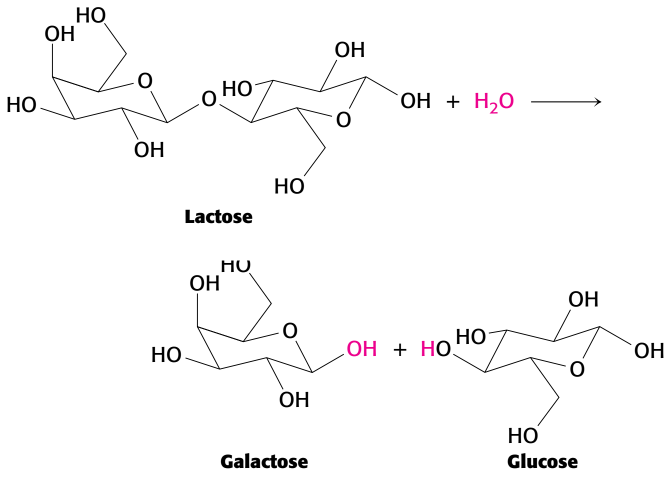
10. *Leaderless.* The mRNA for the λ repressor begins with 5′-AUG-3′, which encodes the methionine residue that begins the protein. What is unusual about this beginning? Would it cause the mRNA to translate efficiently or not?

11. *Quorum count.* Suppose you have a series of compounds that you wish to test for the autoinducer activity in *Vibrio fischeri*. Propose a simple assay, assuming that you can grow *V. fischeri* cultures at low cell densities.

12. *Codon utilization.* There are four codons that encode threonine. Consider the leader sequence in Figure 31.22A. What codons are used and with what frequency?

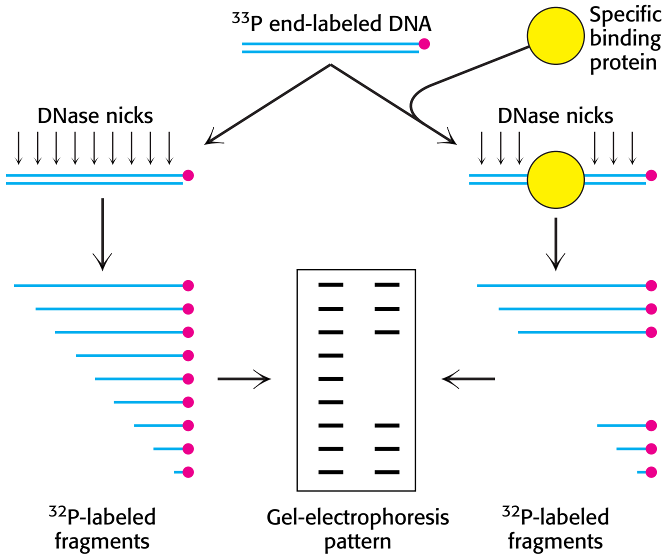
Mechanism Problem

13. *Follow the stereochemistry.* The hydrolysis of lactose is catalyzed by β-galactosidase. Does the overall reaction proceed with retention or inversion of configuration? Given that each step likely proceeds with inversion of configuration, what does the overall change in stereochemistry suggest about the mechanism? A key residue in the reaction has been identified to be Glu 537. Propose an overall mechanism for the hydrolysis of lactose.

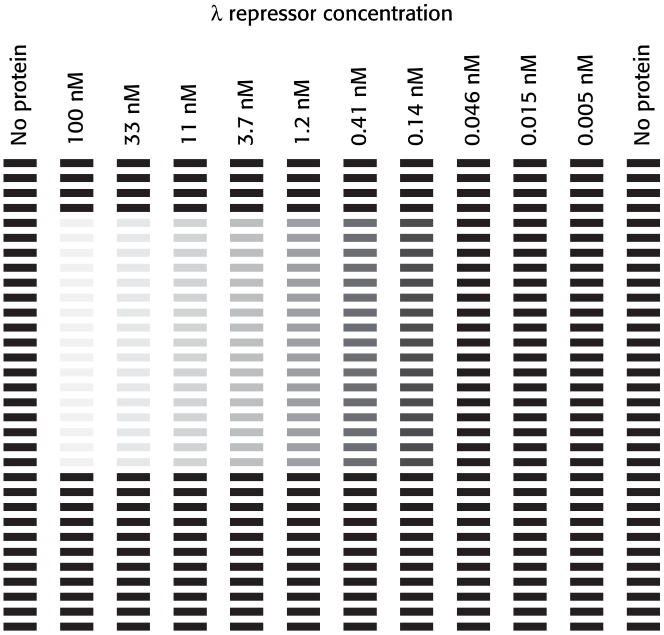


Data Interpretation Problem

14. *Leaving tracks.* A powerful method for examining protein–DNA interactions is called DNA footprinting. In this method, a DNA fragment containing a potential binding site is radiolabeled on one end. The labeled DNA is then treated with a DNA-cleaving agent such as DNase I such that each DNA molecule within the population is cut only once. The same cleavage process is carried out in the presence of the DNA-binding protein. The bound protein protects some sites within the DNA from cleavage. The patterns of DNA fragments in the cleaved pool of DNA molecules are then examined by electrophoresis followed by autoradiography (a).


^{33}P end-labeled DNA
Specific binding protein
DNase nicks
DNase nicks
^{32}P-labeled fragments
Gel-electrophoresis pattern
^{32}P-labeled fragments

This method is applied to a DNA fragment containing a single binding site for the λ repressor in the presence of different concentrations of the λ repressor. The results are shown below (b):

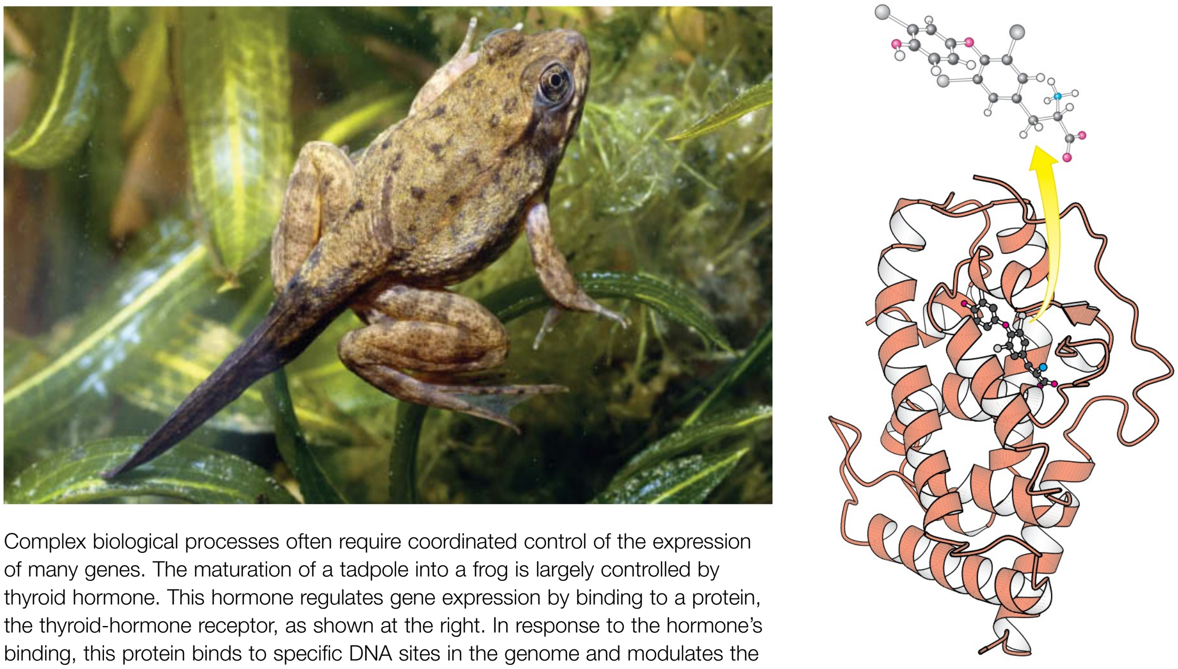


Estimate the dissociation constant for the λ repressor–DNA complex and the standard free energy of binding.

CHAPTER

32

The Control of Gene Expression in Eukaryotes



Complex biological processes often require coordinated control of the expression of many genes. The maturation of a tadpole into a frog is largely controlled by thyroid hormone. This hormone regulates gene expression by binding to a protein, the thyroid-hormone receptor, as shown at the right. In response to the hormone's binding, this protein binds to specific DNA sites in the genome and modulates the expression of nearby genes. [(Left) Gary Meszaros/Getty Images.]

Many of the most important and intriguing features in modern biology and medicine, such as the pathways crucial for the development of multicellular organisms, the changes that distinguish normal cells and cancer cells, and the evolutionary changes leading to new species, entail networks of gene-regulatory pathways. Gene regulation in eukaryotes is significantly more complex than in prokaryotes in several ways. First, the genomes being regulated are significantly larger. The *E. coli* genome consists of a single, circular chromosome containing 4.6 megabases (Mb). This genome encodes approximately 2000 proteins. In comparison, one of the simplest eukaryotes, *Saccharomyces cerevisiae* (baker's yeast), contains 16 chromosomes ranging in size from 0.2 to 2.2 Mb (Figure 32.1). The yeast genome totals 12 Mb and encodes approximately 6000 proteins. The genome within a human cell contains 23 pairs of chromosomes ranging in size from 50 to 250 Mb. Approximately 21,000 genes are present within the 3000 Mb of human DNA.

Second, whereas prokaryotic genomic DNA is relatively accessible, eukaryotic DNA is packaged into chromatin, a complex between the DNA and a special set of proteins (Figure 32.2). Although the principles for the construction of chromatin are relatively simple, the chromatin structure for a complete genome is quite complex. Importantly, in a given eukaryotic cell, some genes and their associated regulatory regions are relatively accessible

OUTLINE

32.1 Eukaryotic DNA Is Organized into Chromatin

32.2 Transcription Factors Bind DNA and Regulate Transcription Initiation

32.3 The Control of Gene Expression Can Require Chromatin Remodeling

32.4 Eukaryotic Gene Expression Can Be Controlled at Posttranscriptional Levels

Megabase

A length of DNA consisting of 10^6 base pairs (if double stranded) or 10^6 bases (if single stranded).

1 Mb = 10^3 kb = 10^6 bases

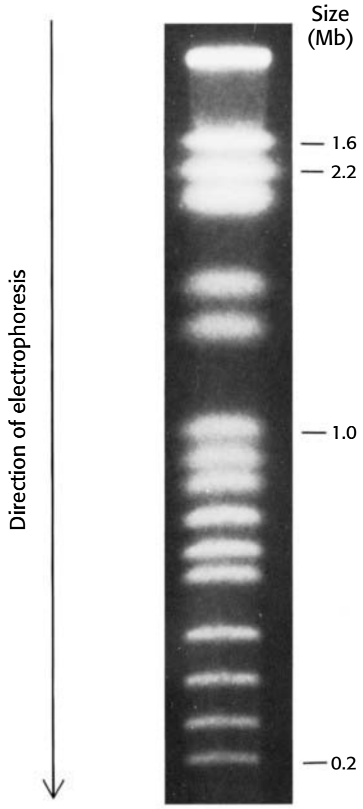


FIGURE 32.1 Yeast chromosomes. Pulsed-field electrophoresis allows the separation of 16 yeast chromosomes. [From G. Chu, D. Vollrath, and R. W. Davis, *Science* 234:1582–1585, 1986.]

for transcription and regulation, whereas other genes are tightly packaged and are thus rendered inactive. Eukaryotic gene regulation frequently requires the manipulation of chromatin structure.

A manifestation of this complexity is the presence of many different *cell types* in most eukaryotes. A liver cell, a pancreatic cell, and an embryonic stem cell contain the same DNA sequences, yet the subset of genes highly expressed in cells from the pancreas, which secretes digestive enzymes, differs markedly from the subset highly expressed in the liver, the site of lipid transport and energy transduction. Embryonic stem cells do not express any subset of genes at high levels; the most highly expressed genes are "housekeeping" genes involved in the cytoskeleton and processes such as translation (Table 32.1). The existence of stable cell types is due to differences in the *epigenome,* differences in chromatin structure, and covalent modifications of the DNA, not in the DNA sequence itself. Thus, different cell types share the same genome (DNA sequence) but differ in their epigenomes, the packaging and modification of this genome.

In addition, eukaryotic genes are not generally organized into operons. Instead, genes that encode proteins for steps within a given pathway are often spread widely across the genome. This characteristic requires that other mechanisms function to regulate genes in a coordinated way.

Despite these differences, some aspects of gene regulation in eukaryotes are quite similar to those in prokaryotes. In particular, activator and repressor proteins that recognize specific DNA sequences are central to many gene-regulatory processes. In this chapter, we will focus first on chromatin structure. We will then turn to transcription factors—DNA-binding proteins similar in many ways to the prokaryotic proteins that we encountered in the preceding chapter. Eukaryotic transcription factors can act directly by interacting with the transcriptional machinery or indirectly by influencing chromatin structure. Finally, we examine selected

TABLE 32.1 Highly expressed protein-encoding genes of the pancreas, liver, and embryonic stem cells (as percentage of total mRNA pool)

Rank	Proteins expressed in pancreas	%	Proteins expressed in liver	%	Proteins expressed in stem cells	%
1	Procarboxypeptidase A1	7.6	Albumin	3.5	Glyceraldehyde-3-phosphate dehydrogenase	0.7
2	Pancreatic trypsinogen 2	5.5	Apolipoprotein A-I	2.8	Translation elongation factor 1 $\alpha 1$	0.6
3	Chymotrypsinogen	4.4	Apolipoprotein C-I	2.5	Tubulin α	0.5
4	Pancreatic trypsin 1	3.7	Apolipoprotein C-III	2.1	Translationally controlled tumor protein	0.5
5	Elastase IIIB	2.4	ATPase 6/8	1.5	Cyclophilin A	0.4
6	Protease E	1.9	Cytochrome oxidase 3	1.1	Cofilin	0.4
7	Pancreatic lipase	1.9	Cytochrome oxidase 2	1.1	Nucleophosmin	0.3
8	Procarboxypeptidase B	1.7	α_1-Antitrypsin	1.0	Connexin 43	0.3
9	Pancreatic amylase	1.7	Cytochrome oxidase 1	0.9	Phosphoglycerate mutase	0.2
10	Bile-salt-stimulated lipase	1.4	Apolipoprotein E	0.9	Translation elongation factor 1 $\beta 2$	0.2

Sources: Data for pancreas from V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler, *Science* 270:484–487, 1995. Data for liver from T. Yamashita, S. Hashimoto, S. Kaneko, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, and K. Matsushima, *Biochem. Biophys. Res. Commun.* 269:110–116, 2000. Data for stem cells from M. Richards, S. P. Tan, J. H. Tan, W. K. Chan, and A. Bongso, *Stem Cells* 22:51–64, 2004.

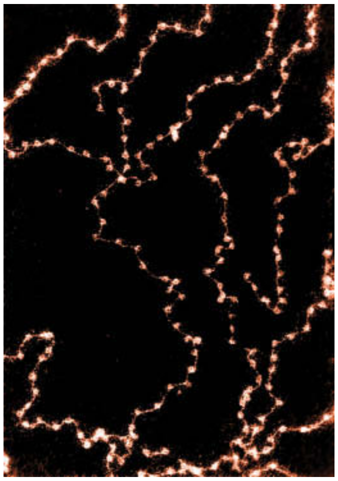


FIGURE 32.2 Chromatin structure. An electron micrograph of chromatin showing its "beads on a string" character. The beads correspond to DNA complexes with specific proteins. [Don W. Fawcett/Science Source.]

posttranscriptional gene-regulatory mechanisms, including those based on microRNAs, an important class of gene-regulatory molecules discovered in recent years.

32.1 Eukaryotic DNA Is Organized into Chromatin

Eukaryotic DNA is tightly bound to a group of small basic proteins called *histones.* In fact, histones constitute half the mass of a eukaryotic chromosome. The entire complex of a cell's DNA and associated protein is called *chromatin.* Chromatin compacts and organizes eukaryotic DNA and its presence has dramatic consequences for gene regulation.

Nucleosomes are complexes of DNA and histones

Chromatin is made up of repeating units, each containing 200 bp of DNA and two copies each of four histone proteins H2A, H2B, H3, and H4. Histones have strikingly basic properties because a quarter of the residues in each histone are either arginine or lysine, positively charged amino acids that strongly interact with the negatively charged DNA. The protein complex is called the *histone octamer*. The repeating units of the histone octamer and the associated DNA are known as *nucleosomes*. Chromatin viewed with the electron microscope has the appearance of beads on a string (Figure 32.2); each bead has a diameter of approximately 100 Å. Partial digestion of chromatin with DNase yields the isolated beads. These particles consist of fragments of DNA about 200 bp in length bound to the histone octamer. More-extensive digestion yields a shorter DNA fragment of 145 bp bound to the octamer. The smaller complex formed by the histone octamer and the 145-bp DNA fragment is the *nucleosome core particle*. The DNA connecting core particles in undigested chromatin is called linker DNA. Histone H1 binds, in part, to the linker DNA.

DNA wraps around histone octamers to form nucleosomes

The overall structure of the nucleosome was revealed through electron microscopic and x-ray crystallographic studies pioneered by Aaron Klug and his colleagues. More recently, the three-dimensional structure of a reconstituted nucleosome core particle (Figure 32.3) was determined to

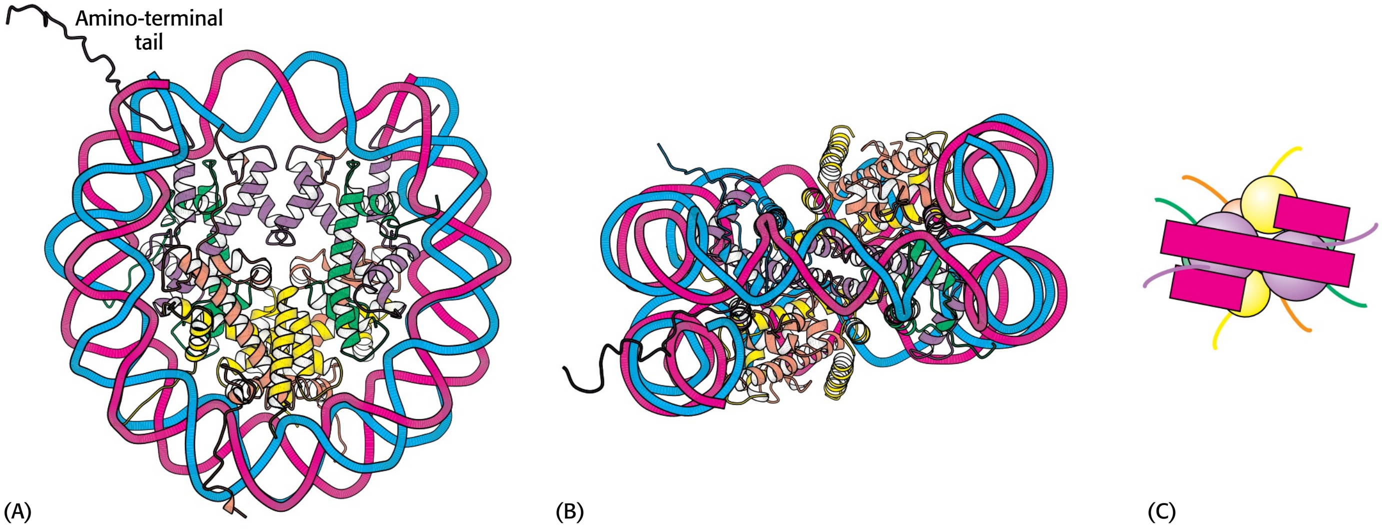


FIGURE 32.3 Nucleosome core particle. The structure consists of a core of eight histone proteins surrounded by DNA. (A) A view showing the DNA wrapping around the histone core. (B) A 90-degree rotation of the view in part (A). *Notice* that the DNA forms a left-handed superhelix as it wraps around the core. (C) A schematic view. [Drawn from 1AOI.pdb.]

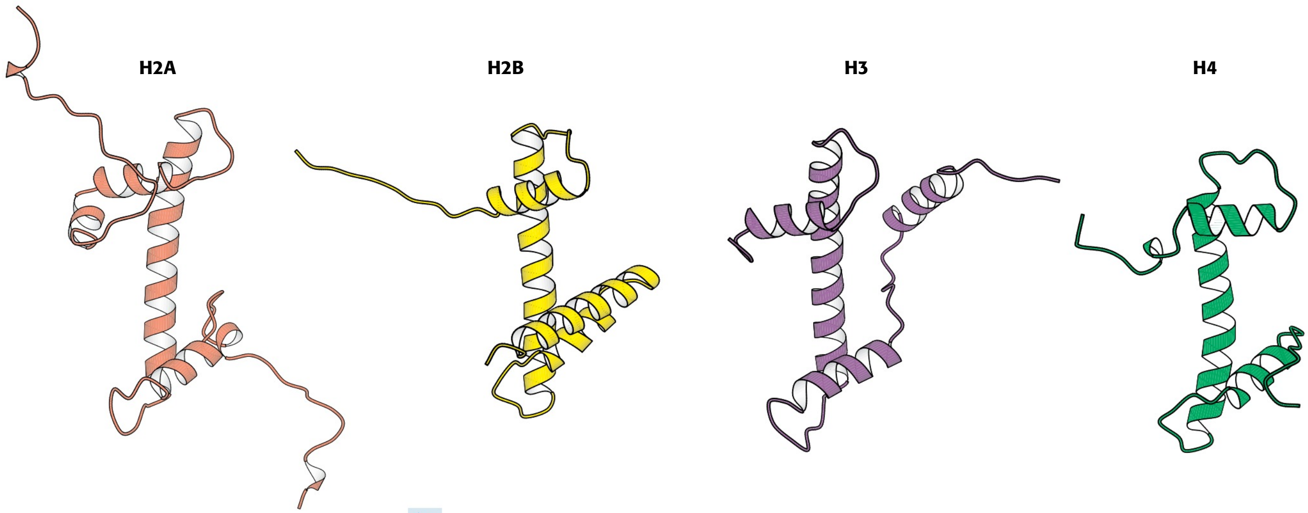


FIGURE 32.4 Homologous histones. Histones H2A, H2B, H3, and H4 adopt a similar three-dimensional structure as a consequence of common ancestry. Some parts of the tails at the termini of the proteins are not shown. [Drawn from 1AOI.pdb.]

higher resolution by x-ray diffraction methods. The four types of histone that make up the protein core are homologous and similar in structure (Figure 32.4). The eight histones in the core are arranged into a $(H3)_2(H4)_2$ tetramer and a pair of H2A–H2B dimers. The tetramer and dimers come together to form a left-handed superhelical ramp around which the DNA wraps. In addition, each histone has an amino-terminal tail that extends out from the core structure. These tails are flexible and contain a number of lysine and arginine residues. As we shall see, *covalent modifications of these tails play an essential role in regulating gene expression.*

The DNA forms a left-handed superhelix as it wraps around the outside of the histone octamer. The protein core forms contacts with the inner surface of the DNA superhelix at many points, particularly along the phosphodiester backbone and the minor groove. Nucleosomes will form on almost all DNA sites, although some sequences are preferred because the dinucleotide steps are properly spaced to favor bending around the histone core. A histone with a different structure from that of the others, called histone H1, seals off the nucleosome at the location at which the linker DNA enters and leaves. The amino acid sequences of histones, including their amino-terminal tails, are remarkably conserved from yeast to human beings.

The winding of DNA around the nucleosome core contributes to the packing of DNA by decreasing its linear extent. An extended 200-bp stretch of DNA would have a length of about 680 Å. Wrapping this DNA around the histone octamer reduces the length to approximately 100 Å along the long dimension of the nucleosome. Thus the DNA is compacted by a factor of 7. However, human chromosomes in metaphase, which are highly condensed, are compacted by a factor of 10^4. Clearly, the nucleosome is just the first step in DNA compaction. What is the next step? The nucleosomes themselves are arranged in a helical array approximately 360 Å across, forming a series of stacked layers approximately 110 Å apart (Figure 32.5). The folding of these fibers of nucleosomes into loops further compacts DNA.

The wrapping of DNA around the histone octamer as a left-handed helix also stores negative supercoils; if the DNA in a nucleosome is straightened out, the DNA will be underwound. This underwinding is exactly what is needed to separate the two DNA strands during replication and transcription.

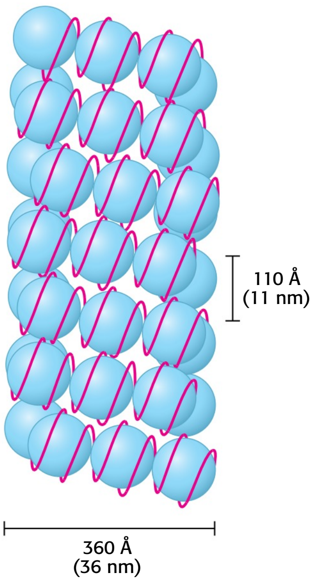


FIGURE 32.5 Higher-order chromatin structure. A proposed model for chromatin arranged in a helical array consisting of six nucleosomes per turn of helix. The DNA double helix (shown in red) is wound around each histone octamer (shown in blue). The actual structure in cells is likely more heterogeneous and dynamic than suggested by this model. [Information from J. T. Finch and A. Klug, *Proc. Natl. Acad. Sci. U. S. A.* 73:1897–1901, 1976.]

32.2 Transcription Factors Bind DNA and Regulate Transcription Initiation

DNA-binding *transcription factors* are key to gene regulation in eukaryotes, just as they are in prokaryotes. However, the roles of eukaryotic transcription factors are different in several ways. First, whereas the DNA-binding sites crucial for the control of gene expression in prokaryotes are usually quite close to promoters, those in eukaryotes can be farther away from promoters and can exert their action at a distance. Second, most prokaryotic genes are regulated by single transcription factors, and multiple genes in a pathway are expressed in a coordinated fashion because such genes are often transcribed as part of a polycistronic mRNA. In eukaryotes, the expression of each gene is typically controlled by multiple transcription factors, and the coordinated expression of different genes depends on having similar transcription-factor-binding sites in each gene in the set. Third, in prokaryotes, transcription factors usually interact directly with RNA polymerase. In eukaryotes, while some transcription factors interact directly with RNA polymerase, many others act more indirectly, interacting with other proteins associated with RNA polymerase or modifying chromatin structure. Let us now examine eukaryotic transcription factors in more detail.

Eukaryotic transcription factors usually consist of several domains. The *DNA-binding domain* binds to regulatory sequences that can either be adjacent to the promoter or at some distance from it. Most commonly, transcription factors include additional domains that help activate transcription. When a transcription factor is bound to the DNA, its *activation domain* promotes transcription by interacting with RNA polymerase II, by interacting with other associated proteins, or by modifying the local structure of chromatin.

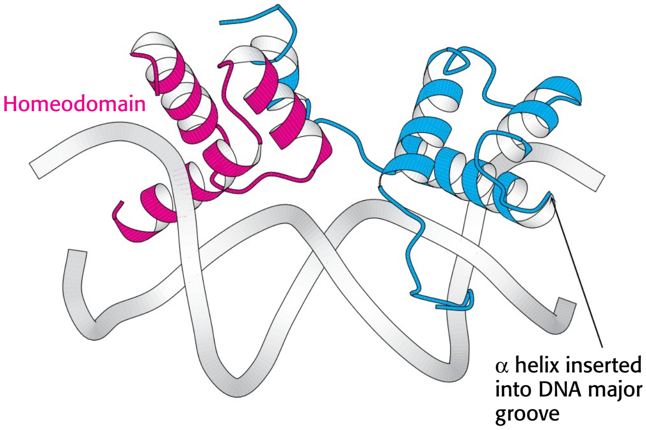


FIGURE 32.6 Homeodomain structure. The structure of a heterodimer formed from two different DNA-binding domains, each based on a homeodomain. *Notice* that each homeodomain has a helix-turn-helix motif with one helix inserted into the major groove of DNA. [Drawn from 1AKH.pdb.]

A range of DNA-binding structures are employed by eukaryotic DNA-binding proteins

The structures of many eukaryotic DNA-binding proteins have been determined and a range of structural motifs have been observed, but we will focus on three that reveal the common features and the diversity of these motifs. The first class of eukaryotic DNA-binding unit that we will consider is the *homeodomain* (Figure 32.6). The structure of the homeodomain and its mode of recognition of DNA are very similar to those of the prokaryotic helix-turn-helix proteins. In eukaryotes, homeodomain proteins often form heterodimeric structures, sometimes with other homeodomain proteins, that recognize asymmetric DNA sequences.

The second class of eukaryotic DNA-binding unit comprises the *basic-leucine zipper* (bZip) *proteins* (Figure 32.7). This DNA-binding unit consists of a pair of long α helices. The first part of each α helix is a basic region that lies in the major groove of the DNA and makes contacts responsible for DNA-site recognition. The second part of each α helix forms a coiled-coil structure with its partner. Because these units are often stabilized by appropriately spaced leucine residues, these structures are often referred to as *leucine zippers*.

The final class of eukaryotic DNA-binding units that we will consider here are the *Cys$_2$His$_2$ zinc-finger domains* (Figure 32.8). A DNA-binding unit of this class comprises tandem sets of small domains, each of which binds a zinc ion through conserved sets of two cysteine and two histidine residues. These structures, often called *zinc-finger domains,* form a string that follows the major

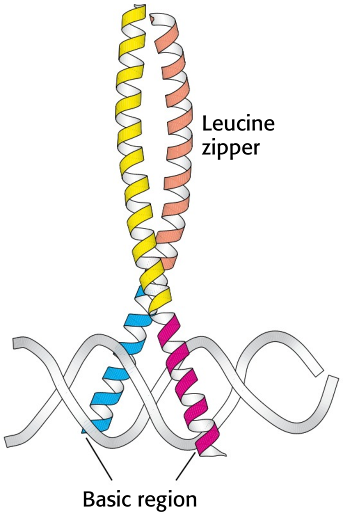


FIGURE 32.7 Basic-leucine zipper. This heterodimer comprises two basic-leucine zipper proteins. *Notice* that the basic region lies in the major groove of DNA. The leucine-zipper part stabilizes the protein dimer. [Drawn from 1FOS.pdb.]

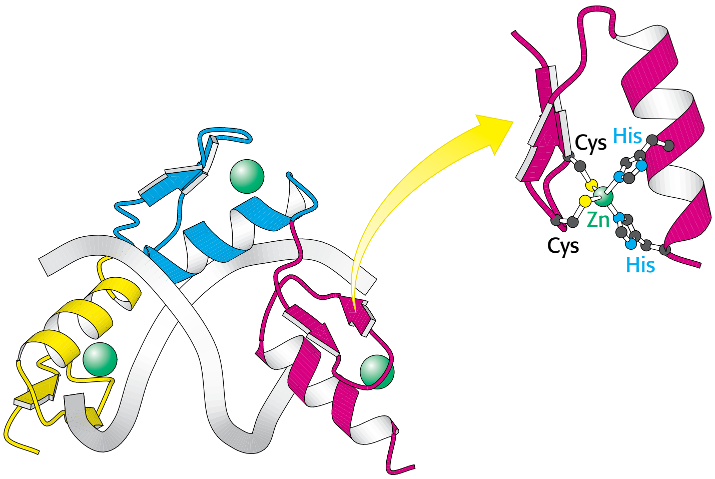


FIGURE 32.8 Zinc-finger domains. A DNA-binding domain comprising three Cys_2His_2 zinc-finger domains (shown in yellow, blue, and red) is shown in a complex with DNA. Each zinc-finger domain is stabilized by a bound zinc ion (shown in green) through interactions with two cysteine residues and two histidine residues. *Notice* how the protein wraps around the DNA in the major groove. [Drawn from 1AAY.pdb.]

groove of DNA. An α helix from each domain makes specific contact with the edges of base pairs within the groove. Some proteins contain arrays of 10 or more zinc-finger domains, potentially enabling them to contact long stretches of DNA. The human genome encodes several hundred proteins that contain zinc-finger domains of this class. We will encounter another class of zinc-based DNA-binding domain when we consider nuclear hormone receptors in Section 32.3.

Activation domains interact with other proteins

The activation domains of transcription factors generally recruit other proteins that promote transcription. In some cases, these activation domains interact directly with RNA polymerase II or closely associated proteins. The activation domains act through intermediary proteins that bridge between the transcription factors and the polymerase. An important target of activators is *mediator,* a complex of 25 to 30 subunits conserved from yeast to human beings, that acts as a bridge between transcription factors and promoter-bound RNA polymerase II (Figure 32.9).

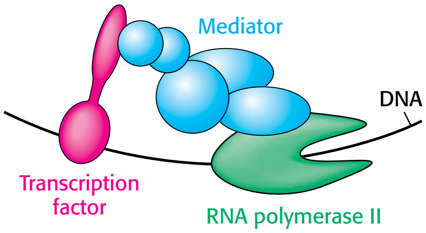


FIGURE 32.9 Mediator. Mediator, a large complex of protein subunits, acts as a bridge between transcription factors bearing activation domains and RNA polymerase II. These interactions help recruit and stabilize RNA polymerase II near specific genes that are then transcribed.

Activation domains are less conserved than DNA-binding domains. In fact, very little sequence similarity has been found. For example, they may be acidic, hydrophobic, glutamine rich, or proline rich. However, they have certain features in common. First, they are often *redundant;* that is, a part of the activation domain can be deleted without loss of function. Second, they are *modular* and can activate transcription when paired with a variety of DNA-binding domains. Third, they can act *synergistically:* two activation domains acting together create a stronger effect than either acting separately.

We have been considering the case in which gene control increases the expression level of a gene. In many cases, the expression of a gene must be decreased by blocking transcription. The agents in such cases are *transcriptional repressors*. Like activators, transcriptional repressors act in many cases by altering chromatin structure.

Multiple transcription factors interact with eukaryotic regulatory regions

The basal transcription complex described in Chapter 29 initiates transcription at a low frequency. Recall that several general transcription factors join with RNA polymerase II to form the basal transcription complex. Additional transcription factors must bind to other sites that can be near the promoter or quite distant for a gene to achieve a higher rate of mRNA synthesis. In contrast with the regulators of prokaryotic transcription, few eukaryotic transcription factors have any effect on their own. Instead, each factor recruits other proteins to build up large complexes that interact with the transcriptional machinery to activate transcription.

A major advantage of this mode of regulation is that a given regulatory protein can have different effects, depending on what other proteins are present in the same cell. This phenomenon, called *combinatorial control,* is crucial to multicellular organisms that have many different cell types. Even in unicellular eukaryotes such as yeast, combinatorial control allows the generation of distinct cell types.

Enhancers can stimulate transcription in specific cell types

Transcription factors can often act even if their binding sites lie at a considerable distance from the promoter. These distant regulatory sites are called *enhancers* (Chapter 29). Enhancers function by serving as binding sites for specific transcription factors. An enhancer is effective only in the specific cell types in which appropriate regulatory proteins are expressed. In many cases, these DNA-binding proteins influence transcription initiation by perturbing the local chromatin structure to expose a gene or its regulatory sites rather than by direct interactions with RNA polymerase. This mechanism accounts for the ability of enhancers to act at a distance.

The properties of enhancers are illustrated by studies of the enhancer controlling the muscle isoform of creatine kinase (Figure 32.10). The results of mutagenesis and other studies revealed the presence of an enhancer located between 1350 and 1050 base pairs upstream of the start site of the gene for this enzyme. Experimentally inserting this enhancer near a gene not normally expressed in muscle cells is sufficient to cause the gene to be expressed at high levels in muscle cells but not in other cells (Figure 32.11).

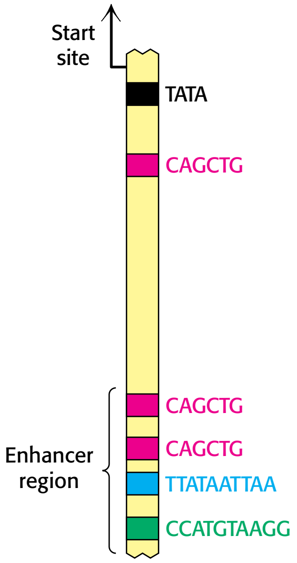


FIGURE 32.10 Enhancer binding sites. A schematic structure for the region 1 kb upstream of the start site for the muscle creatine kinase gene. One binding site of the form 5′-CAGCTG-3′ is present near the TATA box. The enhancer region farther upstream contains two binding sites for the same protein and two additional binding sites for other proteins.

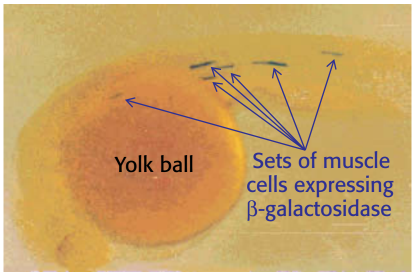


FIGURE 32.11 An experimental demonstration of enhancer function. A promoter for muscle creatine kinase artificially drives the transcription of β-galactosidase in a zebrafish embryo. Only specific sets of muscle cells produce β-galactosidase, as visualized by the formation of the blue product on treatment of the embryo with X-Gal. [From F. Müller, D. W. Williamson, J. Kobolák, L. Gauvry, G. Goldspink, L. Orbán, and N. MacLean, *Mol. Reprod. Dev.* 47:404–412, 1997.]

Induced pluripotent stem cells can be generated by introducing four transcription factors into differentiated cells

An important application illustrating the power of transcription factors is the development of *induced pluripotent stem* (iPS) *cells*. Pluripotent stem cells have the ability to differentiate into many different cell types on appropriate treatment. Previously isolated cells derived from embryos show a very high degree of pluripotency. Over time, researchers identified dozens of genes in embryonic stem cells that contributed to this pluripotency when expressed. In a remarkable experiment reported for mouse cells in 2006 and human cells in 2007, Shinya Yamanaka demonstrated that just four genes out of this entire set could induce pluripotency in already-differentiated skin cells. Yamanaka introduced genes encoding four transcription factors into skin cells called fibroblasts. The fibroblasts de-differentiated into cells that appeared to have characteristics very nearly identical with those of embryonic stem cells (Figure 32.12).

These iPS cells represent powerful research tools and, potentially, a new class of therapeutic agents. The proposed concept is that a sample of a patient's fibroblasts could be readily isolated and converted into iPS cells. These iPS cells could be treated to differentiate into a desired cell type that could then be transplanted into the patient. For example, such an approach might be used to restore a particular class of nerve cells that had been depleted by a neurodegenerative disease. Although the field of iPS cell research is still evolving, it holds great promise as a possible approach to treat any common and difficult-to-treat diseases.

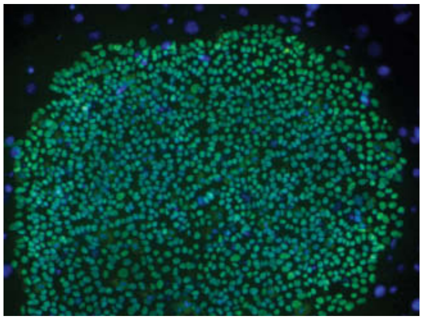


FIGURE 32.12 Induced pluripotent stem cells. A micrograph of human induced pluripotent stem cells stained green for a transcription factor that is characteristic of pluripotent cells. [From K. Takahashi et al., *Cell* 131:861–872, 2007; with permission from Elsevier, courtesy of Shinya Yamanaka, Kyoto University.]

32.3 The Control of Gene Expression Can Require Chromatin Remodeling

Early observations suggested that chromatin structure plays a major role in controlling eukaryotic gene expression. DNA that is densely packaged into chromatin is less susceptible to cleavage by the nonspecific DNA-cleaving enzyme DNase I. Regions adjacent to genes that are being transcribed are more sensitive to cleavage by DNase I than are other sites in the genome, suggesting that the DNA in these regions is less compacted than it is elsewhere and more accessible to proteins. In addition, some sites, usually within 1 kb of the start site of an active gene, are exquisitely sensitive to DNase I and other nucleases. These *hypersensitive sites* correspond to regions that have few nucleosomes or contain nucleosomes in an altered conformation *Hypersensitive sites are cell-type specific and developmentally regulated*. For example, globin genes in the precursors of erythroid cells from 20-hour-old chicken embryos are insensitive to DNase I. However, when hemoglobin synthesis begins at 35 hours, regions adjacent to these genes become highly susceptible to digestion. In tissues such as the brain that produce no hemoglobin, the globin genes remain resistant to DNase I throughout development and into adulthood. These studies suggest that a prerequisite for gene expression is a relaxing of the chromatin structure.

Recent experiments even more clearly revealed the role of chromatin structure in regulating access to DNA-binding sites. Genes required for galactose utilization in yeast are activated by a transcription factor called GAL4, which recognizes DNA-binding sites with two 5′-CGG-3′ sequences on complementary strands separated by 11 base pairs (Figure 32.13). Approximately 4000 potential GAL4 binding sites of the form 5′-CGG(N)$_{11}$CCG-3′ are present in the yeast genome, but only 10 of them regulate genes necessary for galactose metabolism. How is GAL4 targeted to only a small fraction of the potential binding sites? This question is addressed through the use of a technique called *chromatin immunoprecipitation* (ChIP; Figure 32.14). GAL4 is first cross-linked to its DNA-binding sites in chromatin. The DNA is then fragmented into small pieces, and antibodies to GAL4 are used to isolate the chromatin fragments containing GAL4. The cross-linking is reversed, and the DNA is isolated and characterized. The results of these studies reveal that only approximately 10 of the 4000 potential GAL4 sites are occupied by GAL4 when the cells are growing on galactose; more than 99% of the sites appear to be blocked, presumably by the local chromatin structure. Thus, whereas in prokaryotes all sites appear to be equally accessible, chromatin structure

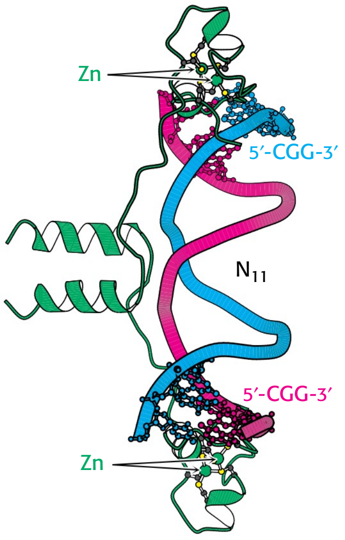


FIGURE 32.13 GAL4 binding sites. The yeast transcription factor GAL4 binds to DNA sequences of the form 5′-CGG(N)$_{11}$CCG-3′. Two zinc-based domains are present in the DNA-binding region of this protein. *Notice* that these domains contact the 5′-CGG-3′ sequences, leaving the center of the site uncontacted. [Drawn from 1D66.pdb.]

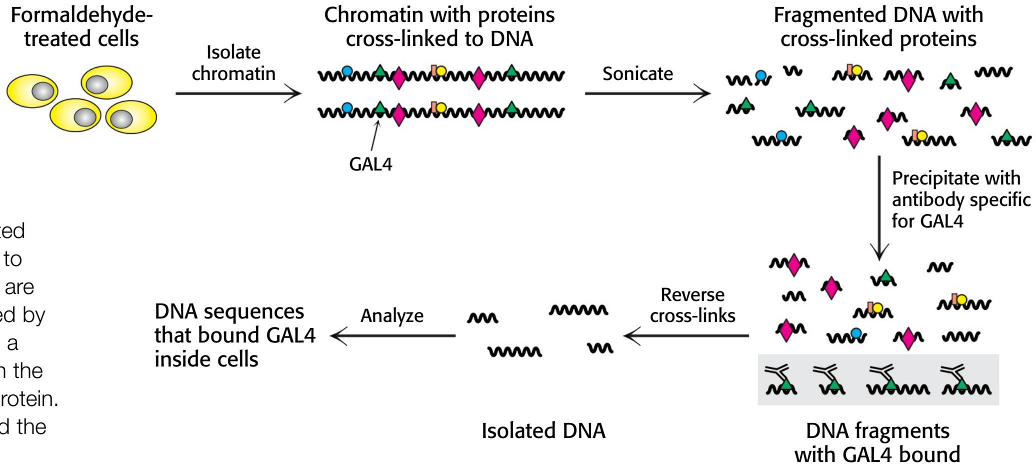
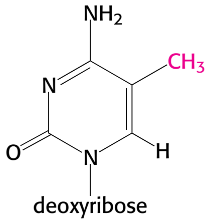

Formaldehyde-treated cells
Isolate chromatin
Chromatin with proteins cross-linked to DNA
GAL4
Sonicate
Fragmented DNA with cross-linked proteins
Precipitate with antibody specific for GAL4
DNA fragments with GAL4 bound
Reverse cross-links
Isolated DNA
Analyze
DNA sequences that bound GAL4 inside cells

FIGURE 32.14 Chromatin immunoprecipitation. Cells or isolated nuclei are treated with formaldehyde to cross-link proteins to DNA. The cells are then lysed and the DNA is fragmented by sonication. DNA fragments bound to a particular protein are isolated through the use of an antibody specific for that protein. The cross-links are then reversed and the DNA fragments are characterized.

shields a large number of the potential binding sites in eukaryotic cells. GAL4 is thereby prevented from binding to sites that are unimportant in galactose metabolism. These lines of evidence and others reveal that chromatin structure is altered in active genes compared with inactive ones.

The methylation of DNA can alter patterns of gene expression

The degree of methylation of DNA provides another mechanism, in addition to packaging with histones, for inhibiting gene expression inappropriate to a specific cell type. Carbon 5 of cytosine can be methylated by specific methyltransferases. About 70% of the 5′-CpG-3′ sequences (where "p" represents the phosphate residue in the DNA backbone) in mammalian genomes are methylated. However, the distribution of these methylated cytosines varies, depending on the cell type. Consider the β-globin gene. In cells that are actively expressing hemoglobin, the region from approximately 1 kb upstream of the start site to approximately 100 bp downstream of the start site is less methylated than the corresponding region in cells that do not express this gene. The relative absence of 5-methylcytosines near the start site is referred to as *hypomethylation.* The methyl group of 5-methylcytosine protrudes into the major groove where it could easily interfere with the binding of proteins that stimulate transcription.

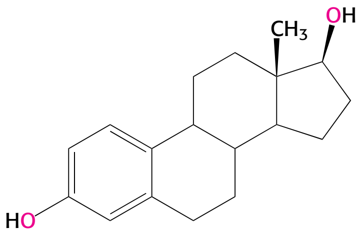


5-Methylcytosine

The distribution of CpG sequences in mammalian genomes is not uniform. Many CpG sequences have been converted into TpG through mutation by the deamination of 5-methylcytosine to thymine. However, sites near the 5′ ends of genes have been maintained because of their role in gene expression. Thus, most genes are found in *CpG islands,* regions of the genome that contain approximately four times as many CpG sequences as does the remainder of the genome.

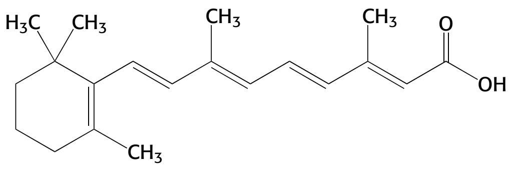
Steroids and related hydrophobic molecules pass through membranes and bind to DNA-binding receptors

We next look at an example that illustrates how transcription factors can stimulate changes in chromatin structure that affect transcription. We will consider in some detail the system that detects and responds to estrogens. Synthesized and released by the ovaries, *estrogens,* such as estradiol, are cholesterol-derived steroid hormones (Section 26.4). They are required for the development of female secondary sex characteristics and, along with progesterone, participate in the ovarian cycle.

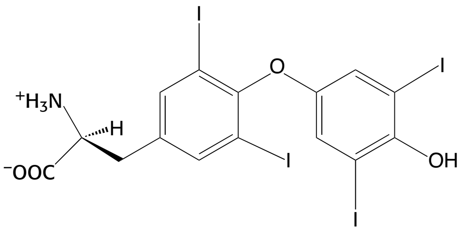


Estradiol
(an estrogen)

Because they are hydrophobic molecules, estrogens easily diffuse across cell membranes. When inside a cell, estrogens bind to highly specific, soluble receptor proteins. Estrogen receptors are members of a large family of proteins that act as receptors for a wide range of hydrophobic molecules, including other steroid hormones, thyroid hormones, and retinoids.



All-*trans*-retinoic acid
(a retinoid)



Thyroxine
(L-3,5,3′,5′-Tetraiodothyronine)
(a thyroid hormone)

The human genome encodes approximately 50 members of this family, often referred to as *nuclear hormone receptors.* The genomes of other

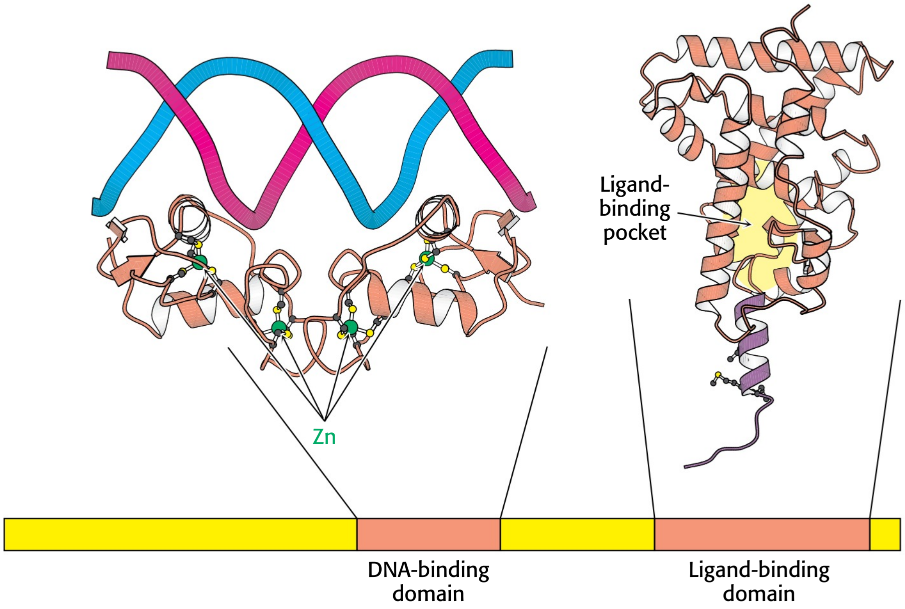


FIGURE 32.15 Structure of two nuclear hormone-receptor domains. Nuclear hormone receptors contain two crucial conserved domains: (1) a DNA-binding domain toward the center of the sequence and (2) a ligand-binding domain toward the carboxyl terminus. The structure of a dimer of the DNA-binding domain bound to DNA is shown, as is one monomer of the normally dimeric ligand-binding domain. [Drawn from 1HCQ and 1LBD.pdb.]

multicellular eukaryotes encode similar numbers of nuclear hormone receptors, although they are absent in yeast.

All these receptors have a similar mode of action. On binding of the signal molecule (called, generically, a *ligand*), the ligand–receptor complex modifies the expression of specific genes by binding to control elements in the DNA. Estrogen receptors bind to specific DNA sites (referred to as *estrogen response elements* or EREs) that contain the consensus sequence 5′-**AGGTCA**NNN**TGACCT**-3′. As expected from the symmetry of this sequence, an estrogen receptor binds to such sites as a dimer.

A comparison of the amino acid sequences of members of this family reveals two highly conserved domains: a DNA-binding domain and a ligand-binding domain (Figure 32.15). The DNA-binding domain lies toward the center of the molecule and consists of a set of zinc-based domains different from the Cys_2His_2 zinc-finger proteins introduced in Section 32.2. These zinc-based domains bind to specific DNA sequences by virtue of an α helix that lies in the major groove in the specific DNA complexes formed by estrogen receptors.

Nuclear hormone receptors regulate transcription by recruiting coactivators to the transcription complex

The second highly conserved domain of the nuclear receptor proteins lies near the carboxyl terminus and is the ligand-binding site. This domain folds into a structure that consists almost entirely of α helices, arranged in three layers. The ligand binds in a hydrophobic pocket that lies in the center of this array of helices (Figure 32.16). This domain changes conformation when it binds its ligand, estrogen. How does ligand binding lead to changes in gene expression? The simplest model would have the binding of ligand alter the DNA-binding properties of the receptor, analogously to the *lac* repressor in prokaryotes. However, experiments with purified nuclear hormone receptors revealed that ligand binding does *not* significantly alter DNA-binding affinity and specificity. Another mechanism is operative.

Investigators sought to determine whether specific proteins might bind to the nuclear hormone receptors only in the presence of ligand. Such searches led to the identification of several related proteins called *coactivators,* such as SRC-1 (*s*teroid *r*eceptor *c*oactivator-1), GRIP-1 (*g*lucocorticoid *r*eceptor *i*nteracting *p*rotein-1), and NcoA-1 (*n*uclear hormone receptor

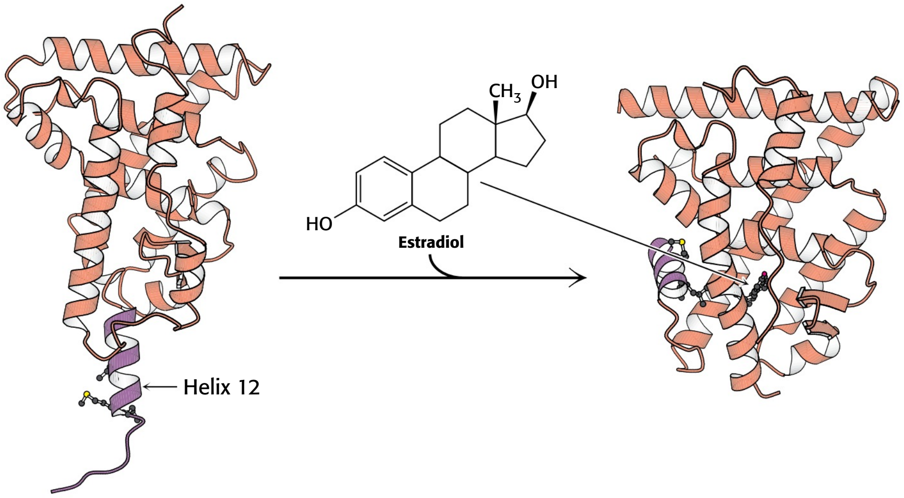
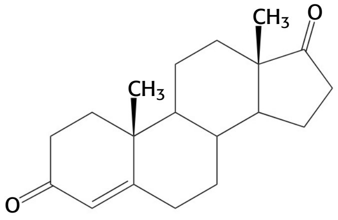


FIGURE 32.16 Ligand binding to nuclear hormone receptor. The ligand lies completely surrounded within a pocket in the ligand-binding domain. *Notice* that the last α helix, helix 12 (shown in purple), folds into a groove on the side of the structure on ligand binding. [Drawn from 1LDB and 1ERE.pdb.]

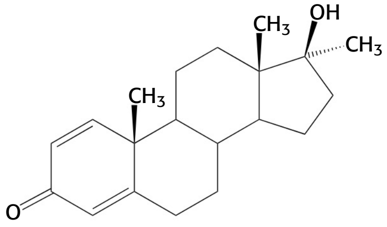
*co*activator-1). These coactivators are referred to as the p160 family because of their size. The binding of ligand to the receptor induces a conformational change that allows the recruitment of a coactivator (Figure 32.17). In many cases, these coactivators are enzymes that catalyze reactions that lead to the modification of chromatin structure.

Steroid-hormone receptors are targets for drugs

Molecules such as estradiol that bind to a receptor and trigger signaling pathways are called *agonists.* Athletes sometimes take natural and synthetic agonists of the androgen receptor, a member of the family of nuclear hormone receptors, because their binding to the androgen receptor stimulates the expression of genes that enhance the development of lean muscle mass.



Androstendione
(a natural androgen)



Dianabol
(methandrostenolone)
(a synthetic androgen)

Referred to as *anabolic steroids,* such compounds used in excess are not without side effects. In men, excessive use leads to a decrease in the secretion of testosterone, to testicular atrophy, and sometimes to breast enlargement

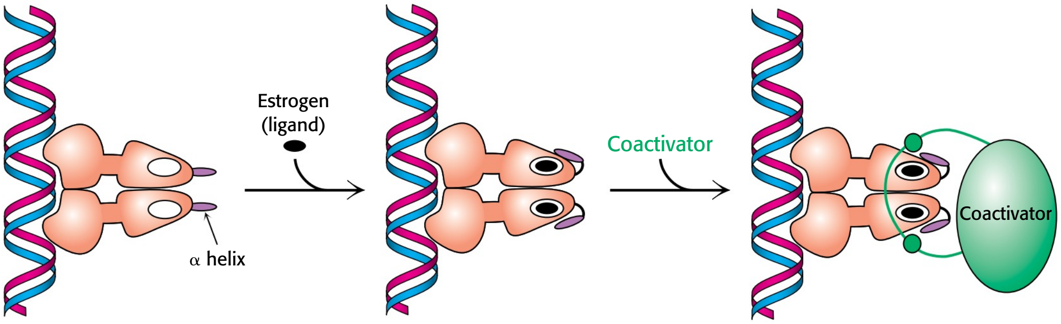


FIGURE 32.17 Coactivator recruitment. The binding of ligand to a nuclear hormone receptor induces a conformational change in the ligand-binding domain. This change in conformation generates favorable sites for the binding of a coactivator.

(gynecomastia) if some of the excess androgen is converted into estrogen. In women, excess testosterone causes a decrease in ovulation and estrogen secretion; it also causes breast regression and growth of facial hair.

Other molecules bind to nuclear hormone receptors but do not effectively trigger signaling pathways. Such compounds are called *antagonists* and are, in many ways, like competitive inhibitors of enzymes. Some important drugs are antagonists that target the estrogen receptor. For example, *tamoxifen* and *raloxifene* are used in the treatment and prevention of breast cancer, because some breast tumors rely on estrogen-mediated pathways for growth. Because some of these compounds have distinct effects on different forms of the estrogen receptor, they are referred to as *selective estrogen receptor modulators* (SERMs).

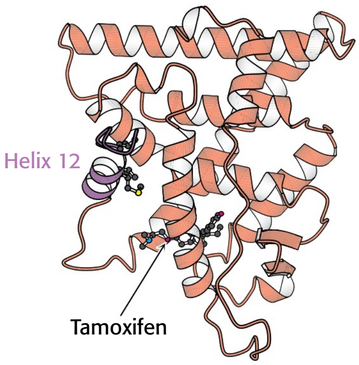
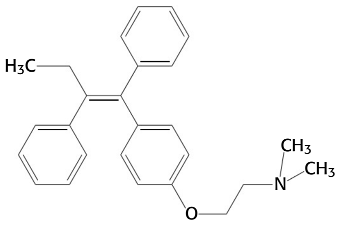
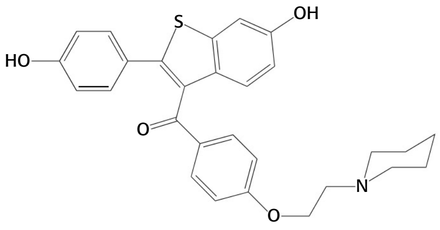


FIGURE 32.18 Estrogen receptor—tamoxifen complex. Tamoxifen binds in the pocket normally occupied by estrogen. However, *notice* that part of the tamoxifen structure extends from this pocket, and so helix 12 cannot pack in its usual position. Instead, this helix blocks the coactivator-binding site. [Drawn from 3ERT.pdb.]



Tamoxifen



Raloxifene

The determination of the structures of complexes between the estrogen receptor and these drugs revealed the basis for their antagonist effect (Figure 32.18). Tamoxifen binds to the same site as estradiol does. However, tamoxifen has a group that extends out of the normal ligand-binding pocket, as do other antagonists. These groups block the normal conformational changes induced by estrogen. Tamoxifen blocks the binding of coactivators and thus inhibits the activation of gene expression.

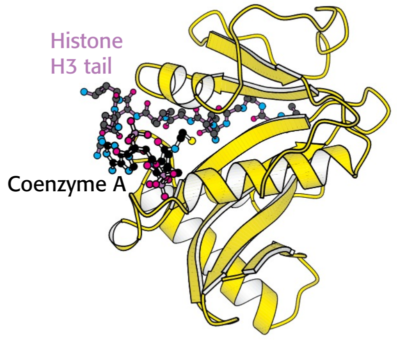
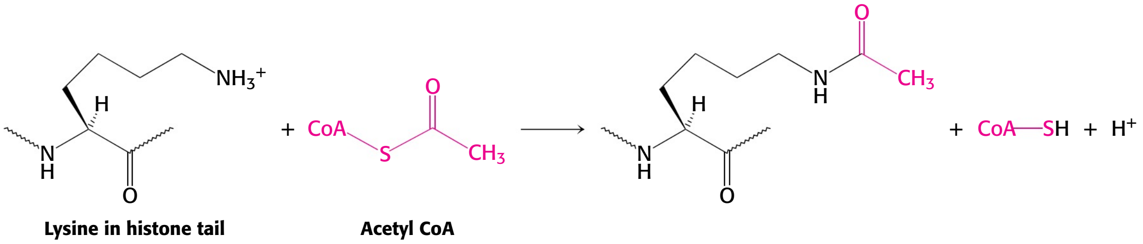


FIGURE 32.19 Structure of histone acetyltransferase. The amino-terminal tail of histone H3 extends into a pocket in which a lysine side chain can accept an acetyl group from acetyl Coenzyme A bound in an adjacent site. [Drawn from 1QSN.pdb.]

Chromatin structure is modulated through covalent modifications of histone tails

We have seen that nuclear receptors respond to signal molecules by recruiting coactivators. Now we can ask, How do coactivators modulate transcriptional activity? *These proteins act to loosen the histone complex from the DNA, exposing additional DNA regions to the transcription machinery.*

Much of the effectiveness of coactivators appears to result from their ability to covalently modify the amino-terminal tails of histones as well as regions on other proteins. Some of the p160 coactivators and the proteins that they recruit catalyze the transfer of acetyl groups from acetyl CoA to specific lysine residues in these amino-terminal tails.



Enzymes that catalyze such reactions are called *histone acetyltransferases* (HATs). The histone tails are readily extended; so they can fit into the HAT active site and become acetylated (Figure 32.19).

What are the consequences of histone acetylation? Lysine bears a positively charged ammonium group at neutral pH. The addition of an acetyl group generates an uncharged amide group. This change dramatically reduces the affinity of the tail for DNA and modestly decreases the affinity of the entire histone complex for DNA, loosening the histone complex from the DNA.

In addition, the acetylated lysine residues interact with a specific *acetyllysine-binding domain* that is present in many proteins that regulate eukaryotic transcription. This domain, termed a *bromodomain,* comprises approximately 110 amino acids that form a four-helix bundle containing a peptide-binding site at one end (Figure 32.20).

Bromodomain-containing proteins are components of two large complexes essential for transcription. One is a complex of more than 10 polypeptides that binds to the *TATA-box-binding protein.* Recall that the TATA-box-binding protein is an essential transcription factor for many genes (Section 29.2). Proteins that bind to the TATA-box-binding protein are called *TAFs* (for *T*ATA-box-binding protein *a*ssociated *f*actors). In particular, TAF1 contains a pair of bromodomains near its carboxyl terminus. The two domains are oriented such that each can bind one of two acetyllysine residues at positions 5 and 12 in the histone H4 tail. Thus, *acetylation of the histone tails provides a mechanism for recruiting other components of the transcriptional machinery.*

Bromodomains are also present in some components of large complexes known as *chromatin-remodeling complexes* or *chromatin-remodeling engines.* These complexes, which also contain domains homologous to those of helicases, utilize the free energy of ATP hydrolysis to shift the positions of nucleosomes along the DNA and to induce other conformational changes in chromatin (Figure 32.21). Histone acetylation can lead to a reorganization of the chromatin structure, potentially exposing binding sites for other factors. *Thus, histone acetylation can activate transcription through a combination of three mechanisms: by reducing the affinity of the histones for DNA, by recruiting other components of the transcriptional machinery, and by initiating the remodeling of the chromatin structure.*

Nuclear hormone receptors also include regions that interact with components of the mediator complex. Thus, two mechanisms of gene regulation can work in concert. Modification of histones and chromatin remodeling can open up regions of chromatin into which the transcription complex can be recruited through protein–protein interactions.

Histone deacetylases contribute to transcriptional repression

Just as in prokaryotes, some changes in a cell's environment lead to the repression of genes that had been active. The modification of histone tails

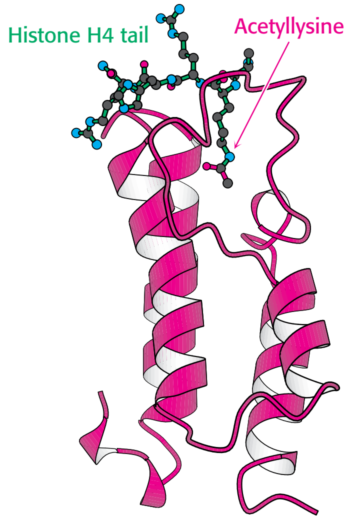
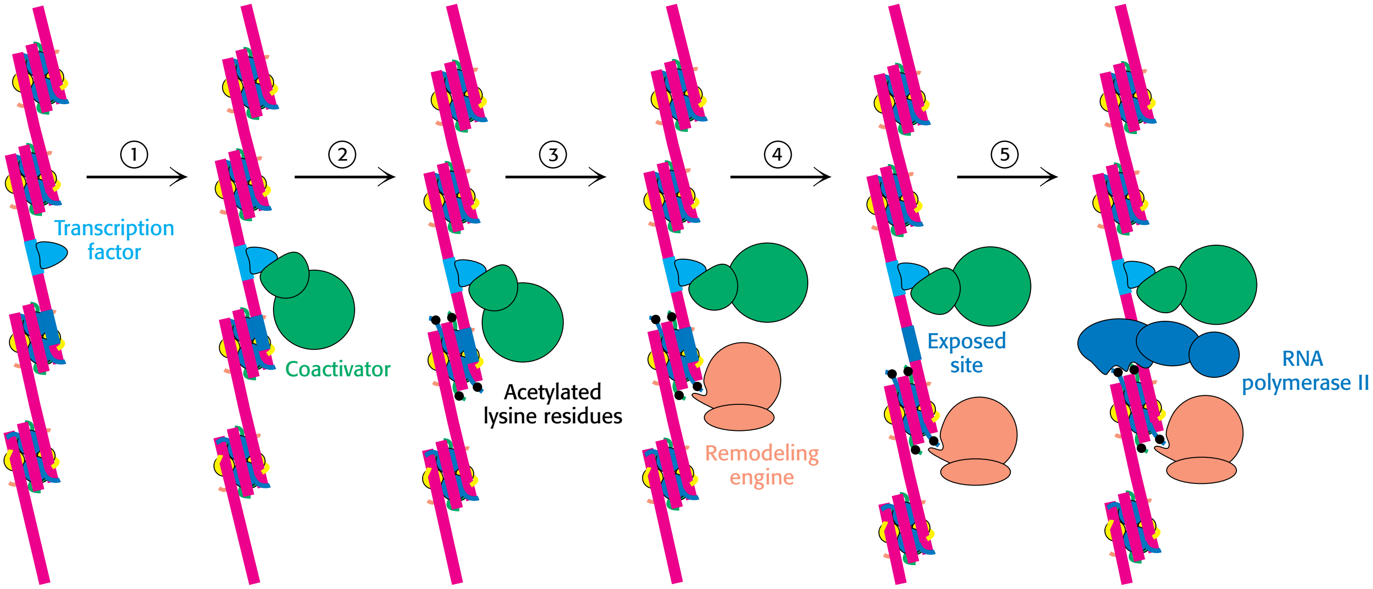


FIGURE 32.20 Structure of a bromodomain. This four-helix-bundle domain binds peptides containing acetyllysine. *Notice* that an acetylated peptide from histone H4 is bound in the structure. [Drawn from 1EGI.pdb.]

FIGURE 32.21 Chromatin remodeling. Eukaryotic gene regulation begins with an activated transcription factor bound to a specific site on DNA. One scheme for the initiation of transcription by RNA polymerase II requires five steps: (1) recruitment of a coactivator, (2) acetylation of lysine residues in the histone tails, (3) binding of a remodeling-engine complex to the acetylated lysine residues, (4) ATP-dependent remodeling of the chromatin structure to expose a binding site for RNA polymerase II or for other factors, and (5) recruitment of RNA polymerase II. Only two subunits are shown for each complex, although the actual complexes are much larger. Other schemes are possible.



again plays an important role. However, in repression, a key reaction appears to be the deacetylation of acetylated lysine, catalyzed by specific *histone deacetylase* enzymes.

In many ways, the acetylation and deacetylation of lysine residues in histone tails (and, likely, in other proteins) is analogous to the phosphorylation and dephosphorylation of serine, threonine, and tyrosine residues in other stages of signaling processes. Like the addition of phosphoryl groups, the addition of acetyl groups can induce conformational changes and generate novel binding sites. Without a means of removal of these groups, however, these signaling switches will become stuck in one position and lose their effectiveness. Like phosphatases, deacetylases help reset the switches.

TABLE 32.2 Selected histone modifications

Modification	Associated with
H4 K8 acetylation	Activation
H3 K14 acetylation	Activation
H3 K27 monomethylation	Activation
H3 K27 trimethylation	Repression
H3 R17 methylation	Activation
H2B S14 phosphorylation	DNA repair
H2B K120 ubiquitination	Activation

Acetylation is not the only modification of histones and other proteins in gene-regulation processes. The methylation of specific lysine and arginine residues also can be important. Some of the more common modifications are shown in Table 32.2.

The elucidation of the roles of these modifications is a very active area of research at present. The relation between various histone modifications and their roles in controlling gene expression is sometimes referred to as "the histone code." Although important generalizations have been discovered, this code clearly is subtle and complicated rather than being a set of hard and fast rules. Nonetheless, the development and application of methods that allow probing of histone modifications on a genome-wide scale in a range of cell types has provided large data sets that can be examined to test and refine models.

32.4 Eukaryotic Gene Expression Can Be Controlled at Posttranscriptional Levels

Just as in prokaryotes, gene expression in eukaryotes can be regulated subsequent to transcription. We shall consider two examples. The first is the regulation of genes taking part in iron metabolism through key features in RNA secondary structure, similar in many ways to prokaryotic posttranscriptional regulation (Section 31.4). The second entails an entirely new mechanism, first glimpsed with the discovery of RNA interference (p. 162). Certain small regulatory RNA molecules allow the regulation of gene expression through interaction with a range of mRNA molecules. Remarkably, this mechanism, discovered only recently, affects the expression of approximately 60% of all human genes.

Genes associated with iron metabolism are translationally regulated in animals

RNA secondary structure plays a role in the regulation of iron metabolism in eukaryotes. Iron is an essential nutrient, required for the synthesis of hemoglobin, cytochromes, and many other proteins. However, excess iron can be quite harmful because, untamed by a suitable protein environment, iron can initiate a range of free-radical reactions that damage proteins, lipids, and nucleic acids. Animals have evolved sophisticated systems for the accumulation of iron in times of scarcity and for the safe storage of excess iron for later use. Key proteins include *transferrin,* a transport protein that carries iron in the serum, *transferrin receptor,* a membrane protein that binds iron-loaded transferrin and initiates its entry into cells, and *ferritin,* an impressively efficient iron-storage protein found primarily in the liver and kidneys. Twenty-four ferritin polypeptides form a nearly spherical shell that encloses as many as 2400 iron atoms, a ratio of one iron atom per amino acid (Figure 32.22).

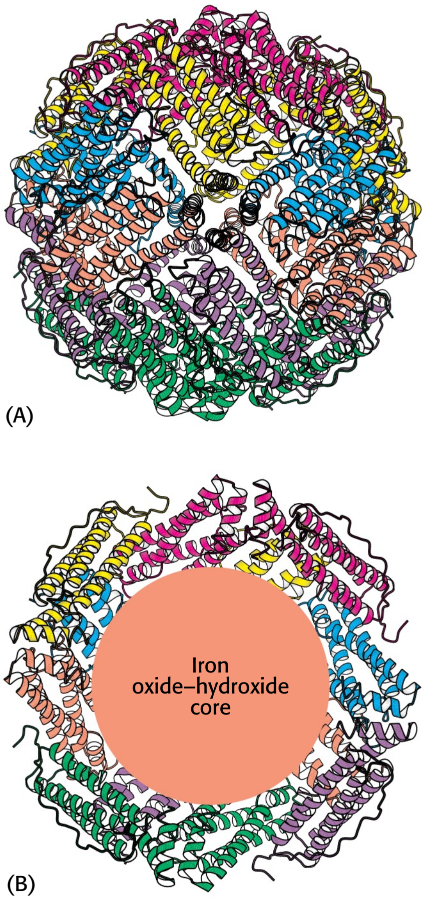


FIGURE 32.22 Structure of ferritin. (A) Twenty-four ferritin polypeptides form a nearly spherical shell. (B) A cutaway view reveals the core that stores iron as an iron oxide–hydroxide complex [Drawn from 1IES.pdb.]

Ferritin and transferrin-receptor expression levels are reciprocally related in their responses to changes in iron levels. When iron is scarce, the amount of

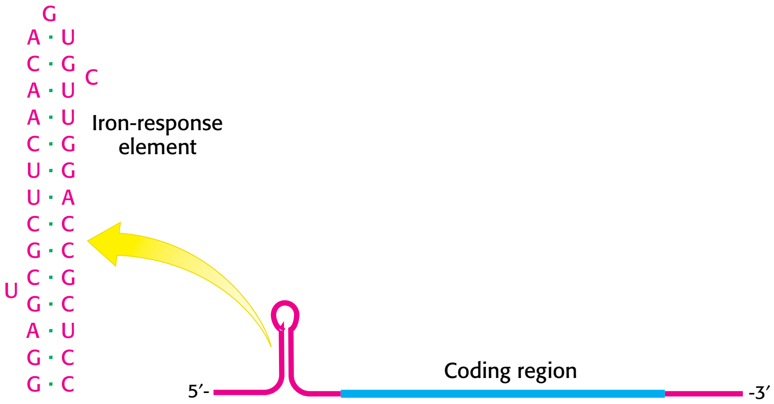


FIGURE 32.23 Iron-response element. Ferritin mRNA includes a stem-loop structure, termed an iron-response element (IRE), in its 5′ untranslated region. The IRE binds a specific protein that blocks the translation of this mRNA under low iron conditions.

transferrin receptor increases and little or no new ferritin is synthesized. Interestingly, the extent of mRNA synthesis for these proteins does not change correspondingly. Instead, regulation takes place at the level of translation.

Consider ferritin first. Ferritin mRNA includes a stem-loop structure termed an *iron-response element* (IRE) in its 5′ untranslated region (Figure 32.23). This stem-loop binds a 90-kDa protein, called an *IRE-binding protein* (IRP), that blocks the initiation of translation as the IRE binds on the 5′ side of the coding region. When the iron level increases, the IRP binds iron as a 4Fe-4S cluster. The IRP bound to iron cannot bind RNA, because the binding sites for iron and RNA substantially overlap. Thus, in the presence of iron, ferritin mRNA is released from the IRP and translated to produce ferritin, which sequesters the excess iron.

An examination of the nucleotide sequence of transferrin-receptor mRNA reveals the presence of several IRE-like regions. However, these regions are located in the 3′ untranslated region rather than in the 5′ untranslated region (Figure 32.24). Under low-iron conditions, IRP binds to these IREs. However, given the location of these binding sites, the transferrin-receptor mRNA can still be translated. What happens when the iron level increases and the IRP no longer binds transferrin-receptor mRNA? Freed from the IRP, transferrin-receptor mRNA is rapidly degraded. Thus, an increase in the cellular iron level leads to the destruction of transferrin-receptor mRNA and, hence, a reduction in the production of transferrin-receptor protein.

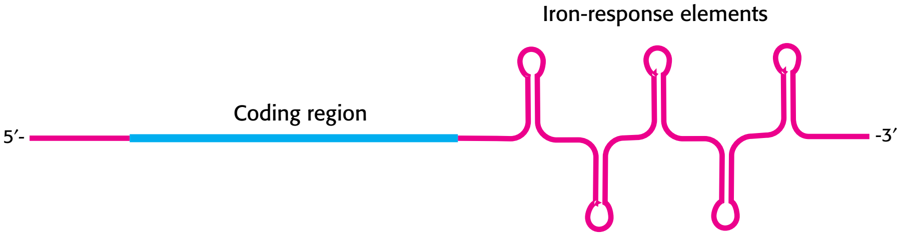


FIGURE 32.24 Transferrin-receptor mRNA. This mRNA has a set of iron-response elements (IREs) in its 3′ untranslated region. The binding of the IRE-binding protein to these elements stabilizes the mRNA but does not interfere with translation.

The purification of the IRP and the cloning of its cDNA were sources of truly remarkable insight into evolution. The IRP was found to be approximately 30% identical in amino acid sequence with the citric acid cycle enzyme aconitase from mitochondria. Further analysis revealed that the IRP is, in fact, an active aconitase enzyme; it is a cytoplasmic aconitase that had been known for a long time, but its function was not well understood (Figure 32.25). The iron–sulfur center at the active site of the IRP is rather unstable, and loss

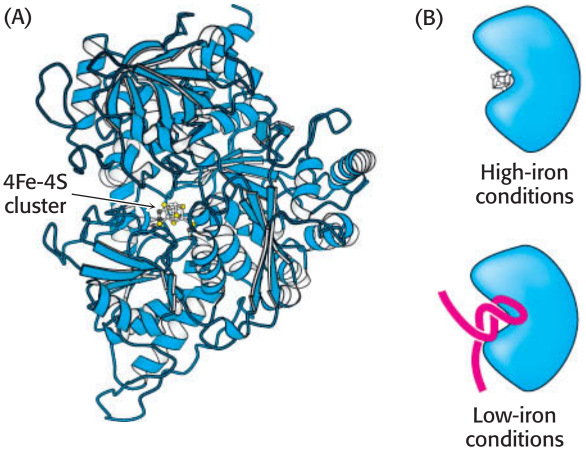


FIGURE 32.25 The IRP is an aconitase. (A) Aconitase contains an unstable 4Fe-4S cluster at its center. (B) Under conditions of low iron, the 4Fe-4S cluster dissociates and appropriate RNA molecules can bind in its place. [Drawn from 1C96.pdb.]

of the iron triggers significant changes in protein conformation. Thus, this protein can serve as an iron-sensing factor.

Other mRNAs, including those taking part in heme synthesis, have been found to contain IREs. Thus, genes encoding proteins required for iron metabolism acquired sequences that, when transcribed, provided binding sites for the iron-sensing protein. An environmental signal—the concentration of iron—controls the translation of proteins required for the metabolism of this metal. Thus, mutations in the untranslated region of mRNAs have been selected for beneficial regulation by iron levels.

Small RNAs regulate the expression of many eukaryotic genes

Genetic studies of development in *C. elegans* revealed that a gene called *lin-4* encodes an RNA molecule 61 nucleotides long that can regulate the expression of certain other genes. The 61-nucleotide *lin-4* RNA does not encode a protein, but rather is cleaved into a 22-nucleotide RNA that possesses the regulatory activity. This discovery was the first view of a large class of regulatory RNA molecules that are now called *microRNAs* or *miRNAs*. The key to the activity of microRNAs and their specificity for particular genes is their ability to form Watson–Crick base-pair-stabilized complexes with the mRNAs of those genes.

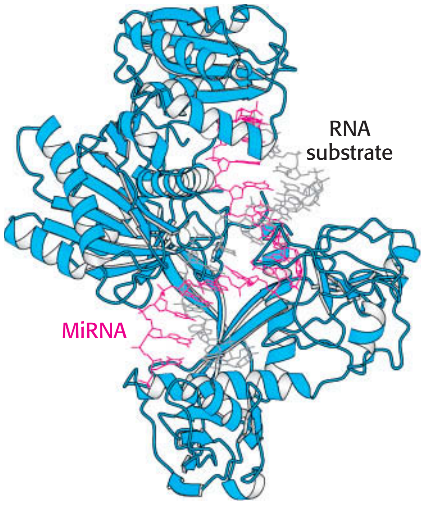


FIGURE 32.26 MicroRNA–Argonaute complex. The miRNA (shown in red) is bound by the Argonaute protein. *Notice* that the miRNA serves as a guide that binds the RNA substrate (shown in gray) through the formation of a double helix. Two magnesium ions are shown in green. [Drawn from 3HK2.pdb.]

These miRNAs do not function on their own. Instead, the miRNAs bind to members of a class of proteins called the *Argonaute* family (Figure 32.26). These Argonaute–miRNA complexes can then bind mRNAs that have sequences that are substantially complementary to the miRNAs. Once bound, such an mRNA can be cleaved by the Argonaute–miRNA complex through the action of a magnesium-based active site. Thus, *the miRNAs serve as guide RNAs that determine the specificity of the Argonaute complex* (Figure 32.27). Cleavage of mRNA by the Argonaute–miRNA complex is similar to the mechanism of RNA interference (p. 162). The small RNA molecules participating in RNAi, however, come from a different source. In RNAi, double-stranded RNAs are cleaved into 21-nucleotide fragments, single-stranded components of which are bound by members of the Argonaute family to form an RNA-induced silencing complex (RISC) that cleaves complementary mRNAs. For miRNAs, the single-stranded RNAs are generated from larger genetically encoded precursors, as described in Chapter 29.

The occurrence of gene regulation by miRNAs was originally thought to be limited to a relatively small number of species. However, subsequent studies have revealed that this mode of gene regulation is nearly ubiquitous in eukaryotes. Indeed, more than 700 miRNAs encoded by the human genome have been identified. Each miRNA can regulate many different genes because many different target sequences are present in each mRNA. An estimated 60% of all human genes are regulated by one or more miRNAs. As one example, consider the human miRNA called miR-206. This miRNA down-regulates the expression of one isoform of the estrogen receptor. In addition, it appears to down-regulate the expression of several different coactivators that interact with the estrogen receptor. Thus, this

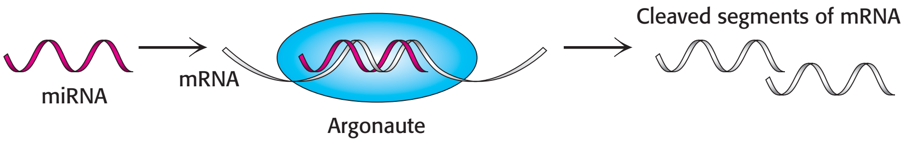


FIGURE 32.27 MicroRNA action. MicroRNAs bind to members of the Argonaute family where they target specific mRNA molecules for cleavage.

miRNA can mute the influence of estrogen by blocking the estrogen-initiated signaling pathway at several different steps.

The microRNA pathway has tremendous implications for the evolution of gene-regulatory pathways. Most of the target sites for miRNAs are present in the $3'$-untranslated regions of mRNAs. These sequences are quite free to mutate because they do not encode proteins and are not required to fold into any specific structures. Thus, in the context of a set of expressed miRNAs, mutations in this region of any particular gene could, in principle, increase or decrease the affinity for one or more miRNAs and, hence, alter the regulation of the gene.

SUMMARY

32.1 Eukaryotic DNA Is Organized into Chromatin

Eukaryotic DNA is tightly bound to basic proteins called histones; the combination is called chromatin. DNA wraps twice around an octamer of core histones to form a nucleosome. The four core histones are homologous and fold into similar structures. Each core histone has an amino-terminal tail rich in lysine and arginine residues. Nucleosomes are the first stage of compaction of eukaryotic DNA. Chromatin blocks access to many potential DNA-binding sites. Changes in chromatin structure play a major role in regulating gene expression.

32.2 Transcription Factors Bind DNA and Regulate Transcription Initiation

Most eukaryotic genes are not expressed unless they are activated by the binding of specific proteins, called transcription factors, to sites on the DNA. These specific DNA-binding proteins interact directly or indirectly with RNA polymerases or their associated proteins. Eukaryotic transcription factors are modular: they consist of separate DNA-binding and activation domains. Important classes of DNA-binding proteins include the homeodomains, the basic-leucine zipper proteins, and Cys$_2$His$_2$ zinc-finger proteins. Each of these classes of proteins uses an α helix to make specific contacts with DNA. Activation domains interact with RNA polymerases or their associated factors or with other protein complexes such as mediator. Enhancers are DNA elements that can modulate gene expression from more than 1000 bp away from the start site of transcription. Enhancers are often specific for certain cell types, depending on which DNA-binding proteins are present. The introduction of genes for a specific set of four transcription factors into fibroblasts can cause these cells to dedifferentiate into induced pluripotent stem cells.

32.3 The Control of Gene Expression Can Require Chromatin Remodeling

Chromatin structure is crucial to the control of gene expression; it is more open near the transcription start sites of actively transcribed genes. Steroids such as estrogens bind to eukaryotic transcription factors called nuclear hormone receptors. These proteins are capable of binding DNA whether or not ligands are bound. The binding of ligands induces a conformational change that allows the recruitment of additional proteins called coactivators. Among the most important functions of coactivators is to catalyze the addition of acetyl groups to lysine residues in the tails of histone proteins. Histone acetylation decreases the affinity of the histones for DNA, making additional genes accessible for transcription. In addition, acetylated histones are targets for proteins containing specific binding units called bromodomains. Bromodomains are components of two classes of large complexes: (1) chromatin-remodeling engines and (2) factors associated with RNA polymerase II. These complexes open up sites on chromatin and initiate transcription.

32.4 Eukaryotic Gene Expression Can Be Controlled at Posttranscriptional Levels

Genes encoding proteins that transport and store iron are regulated at the translational level. Iron-response elements, structures that are present in certain mRNAs, are bound by an IRE-binding protein when this protein is not binding iron. Whether the expression of a gene is stimulated or inhibited in response to changes in the iron status of a cell depends on the location of the IRE within the mRNA. The IRE-binding protein is a cytoplasmic aconitase that loses its iron–sulfur center under low-iron conditions. MicroRNAs are specialized RNA molecules encoded as parts of larger RNA precursors. They bind to proteins of the Argonaute family; the bound miRNAs function as guides that help bind specific mRNA molecules that are then cleaved.

KEY TERMS

cell type (p. 942)
histone (p. 943)
chromatin (p. 943)
nucleosome (p. 943)
nucleosome core particle (p. 943)
transcription factor (p. 945)
homeodomain (p. 945)
basic-leucine zipper (bZip) protein (p. 945)
Cys_2His_2 zinc-finger domain (p. 945)
mediator (p. 946)
combinatorial control (p. 946)
enhancer (p.947)
induced pluripotent stem (iPS) cell (p. 947)
hypersensitive site (p. 948)
chromatin immunoprecipitation (ChIP) (p. 948)
hypomethylation (p. 949)
CpG island (p. 949)
nuclear hormone receptor (p. 949)
estrogen response element (ERE) (p. 950)
coactivator (p. 950)
agonist (p. 951)
anabolic steroid (p. 951)
antagonist (p. 952)
selective estrogen receptor modulator (SERM) (p. 952)
histone acetyltransferase (HAT) (p. 952)
acetyllysine-binding domain (p. 953)
bromodomain (p. 953)
TATA-box-binding protein associated factor (TAF) (p. 953)
chromatin-remodeling complex (p. 953)
histone deacetylase (p. 954)
transferrin (p. 954)
transferrin receptor (p. 954)
ferritin (p. 954)
iron-response element (IRE) (p. 955)
IRE-binding protein (IRP) (p. 955)
microRNA (miRNA) (p. 956)
Argonaute family proteins (p. 956)

PROBLEMS

1. *Charge neutralization.* Given the histone amino acid sequences illustrated below, estimate the charge of a histone octamer at pH 7. Assume that histidine residues are uncharged at this pH. How does this charge compare with the charge on 150 base pairs of DNA?

Histone H2A

MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAAVLEYLTAEILELAGNA

ARDNKKTRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKKTESHHKAKGK

Histone H2B

MPEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDI

FERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK

Histone H3

MARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQR

LVREIAQDFKTDLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

Histone H4

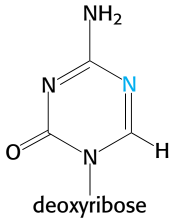
MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDA

VTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG

2. *Chromatin immunoprecipitation.* You have used the technique of chromatin immunoprecipitation to isolate DNA fragments containing a DNA-binding protein of interest. Suppose that you wish to know whether a particular known DNA fragment is present in the isolated mixture. How might you detect its presence?

3. *Around we go.* Assuming that 145 base pairs of DNA wrap around the histone octamer 1.75 times, estimate the radius of the histone octamer. Assume 3.4 Å per base pair and simplify the calculation by assuming that the wrapping is in two rather than three dimensions and neglecting the thickness of the DNA.

4. *Nitrogen substitution.* Growth of mammalian cells in the presence of 5-azacytidine results in the activation of some normally inactive genes. Propose an explanation.



5-Azacytidine

5. *A new domain*. A protein domain that recognizes 5-methylcytosine in the context of double-stranded DNA has been characterized. What role might proteins containing such a domain play in regulating gene expression? Where on a double-stranded DNA molecule would you expect such a domain to bind?

6. *Hybrid receptor*. Through recombinant DNA methods, a modified steroid hormone receptor was prepared that consists of an estrogen receptor with its ligand-binding domain replaced by the ligand-binding domain from the progesterone receptor. Predict the expected responsiveness of gene expression for cells treated with estrogen or with progesterone.

7. *Different modifications*. What is the effect of acetylation of a lysine residue on the charge of a histone protein? Of lysine methylation?

8. *Transformer*. The following amino acid sequence of one of the four transcription factors is used to generate iPS cells:

HTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQKCDRAFSRSDHLALHMKRHF

This transcription factor belongs to one of the three structural classes discussed in Section 32.2. Identify the class.

9. *Coverage*. What percentage of the DNA sites in yeast are accessible, assuming that the fraction of sites observed for GAL4 is typical? To how many base pairs of the 12-Mb yeast genome does this percentage correspond?

10. *Count the methyl groups*. Examination of the histone modifications of a gene reveals an abundance of histone H3 with lysine 27 modified with a single methyl group. Does this suggest that this gene is activated or repressed? How would your answer change if many lysine 27 residues were modified with three methyl groups each?

11. *Iron regulation*. What effect would you expect from the addition of an IRE to the 5′ end of a gene that is not normally regulated by iron levels? To the 3′ end?

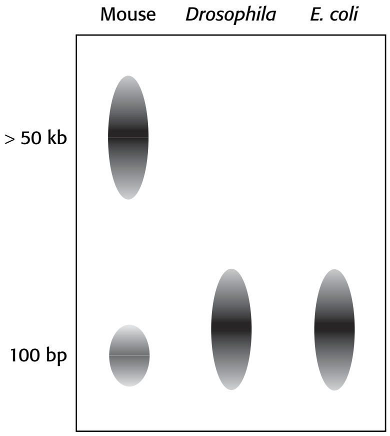
12. *Predicting microRNA regulation*. Suppose that you have identified an miRNA that has the sequence 5′-GCCUAGCCUUAGCAUUGAUUGG-3′. Propose a strategy for identifying mRNA that might be regulated by this miRNA, given the sequences of all mRNAs encoded by the human genome.

Mechanism Problem

13. *Acetyltransferases*. Propose a mechanism for the transfer of an acetyl group from acetyl CoA to the amino group of lysine.

Data Interpretation Problem

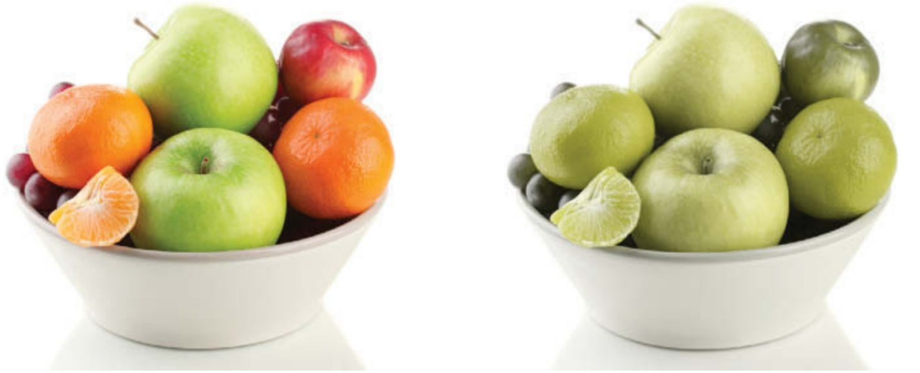
14. *Limited restriction*. The restriction enzyme HpaII is a powerful tool for analyzing DNA methylation. This enzyme cleaves sites of the form 5′-CCGG-3′ but will not cleave such sites if the DNA is methylated on any of the cytosine residues. Genomic DNA from different organisms is treated with HpaII and the results are analyzed by gel electrophoresis (see the adjoining patterns). Provide an explanation for the observed patterns.

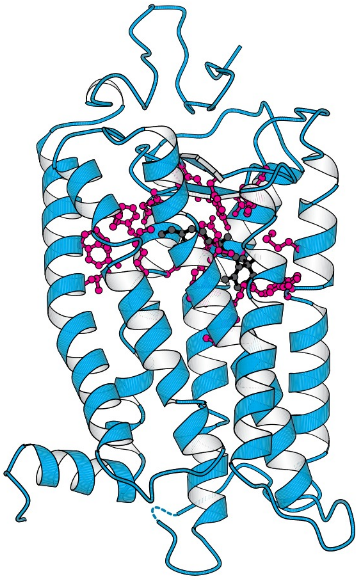


CHAPTER

33

Sensory Systems





Color perception requires specific photoreceptors. The photoreceptor rhodopsin (right), which absorbs light in the process of vision, consists of the protein opsin and a bound vitamin A derivative, retinal. The amino acids (shown in red) that surround the retinal determine the color of light that is most efficiently absorbed. A person lacking a light-absorbing photoreceptor for the color green will see a bowl of fruits of many colors (left) as similar shades of yellow-green (middle). [From Africa Studio/Shutterstock (left), modified by Jeremy Berg using Coblis (www.color-blindness.com/coblis-color-blindness-simulator/) (middle).]

OUTLINE

33.1 A Wide Variety of Organic Compounds Are Detected by Olfaction

33.2 Taste Is a Combination of Senses That Function by Different Mechanisms

33.3 Photoreceptor Molecules in the Eye Detect Visible Light

33.4 Hearing Depends on the Speedy Detection of Mechanical Stimuli

33.5 Touch Includes the Sensing of Pressure, Temperature, and Other Factors

Our senses provide us with means for detecting a diverse set of external signals, often with incredible sensitivity and specificity. For example, when fully adapted to a darkened room, our eyes allow us to sense very low levels of light, down to a limit of less than 10 photons. With more light, we are able to distinguish millions of colors. Through our senses of smell and taste, we are able to detect thousands of chemicals in our environment and sort them into categories: pleasant or unpleasant? healthful or toxic? Finally, we can perceive mechanical stimuli in the air and around us through our senses of hearing and touch.

How do our sensory systems work? How are the initial stimuli detected? How are these initial biochemical events transformed into perceptions and experiences? We have already encountered systems that sense and respond to chemical signals—namely, receptors that bind to growth factors and hormones. Our knowledge of these receptors and their associated signal-transduction pathways provides us with concepts and tools for unraveling some of the workings of sensory systems. For example, 7TM receptors (seven-transmembrane receptors; Section 14.1) play key roles in olfaction, taste, and vision. Ion channels that are sensitive to mechanical stress are essential for hearing and touch.

In this chapter, we focus on the five major sensory systems found in human beings and other mammals: olfaction (the sense of smell—i.e., the detection of small molecules in the air), taste, or gustation (the detection of selected organic compounds and ions by the tongue), vision (the detection

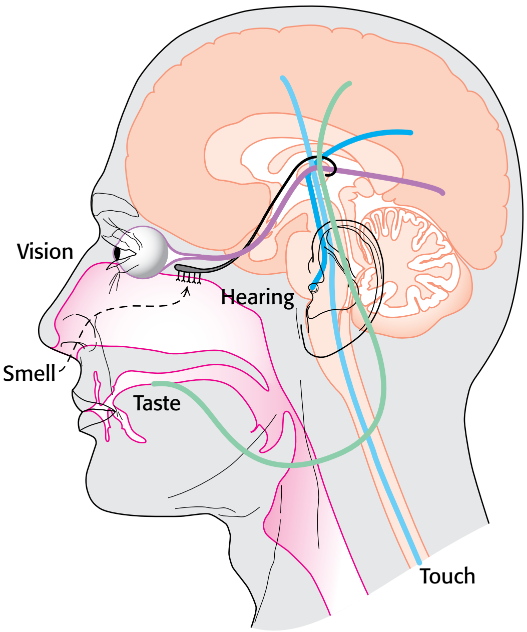
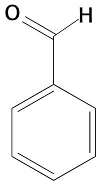


FIGURE 33.1 Sensory connections to the brain. Sensory nerves connect sensory organs to the brain and spinal cord.

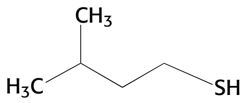
of light), hearing (the detection of sound, or pressure waves in the air), and touch (the detection of changes in pressure, temperature, and other factors by the skin). Each of these primary sensory systems contains specialized sensory neurons that transmit nerve impulses to the central nervous system (Figure 33.1). In the central nervous system, these signals are processed and combined with other information to yield a perception that may trigger a change in behavior. By these means, our senses allow us to detect changes in our environments and to adjust our behavior appropriately.

33.1 A Wide Variety of Organic Compounds Are Detected by Olfaction

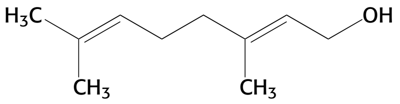
Human beings can detect and distinguish thousands of different compounds by smell, often with considerable sensitivity and specificity. Most odorants are small organic compounds with sufficient volatility that they can be carried as vapors into the nose. For example, a major component responsible for the odor of almonds is the simple aromatic compound benzaldehyde, whereas the sulfhydryl compound 3-methylbutane-1-thiol is a major component of the odor of skunks.



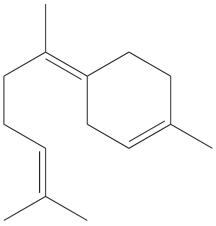
Benzaldehyde
(Almond)



3-Methylbutane-1-thiol
(Skunk)

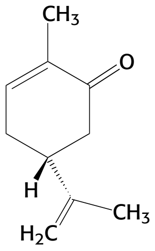


Geraniol
(Rose)

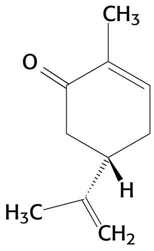


Zingiberene
(Ginger)

What properties of these molecules are responsible for their odors? First, *the shape of the molecule rather than its other physical properties is crucial*. We can most clearly see the importance of shape by comparing molecules such as those responsible for the odors of spearmint and caraway.



***R*-Carvone**
(Spearmint)



***S*-Carvone**
(Caraway)

These compounds are identical in essentially all physical properties such as hydrophobicity because they are exact mirror images of one another. Thus, the odor produced by an odorant depends not on a physical property but on the compound's interaction with a specific binding surface, most likely a protein receptor. Second, some human beings (and other animals) suffer from *specific anosmias;* that is, they are incapable of smelling specific compounds even though their olfactory systems are otherwise normal. Such anosmias are often inherited. These observations suggest that mutations in individual receptor genes lead to the loss of the ability to detect a small subset of compounds.

Olfaction is mediated by an enormous family of seven-transmembrane-helix receptors

Odorants are detected in a specific region of the nose, called the *main olfactory epithelium,* that lies at the top of the nasal cavity (Figure 33.2). Approximately 1 million sensory neurons line the surface of this region. Cilia containing the odorant-binding protein receptors project from these neurons into the mucous lining of the nasal cavity.

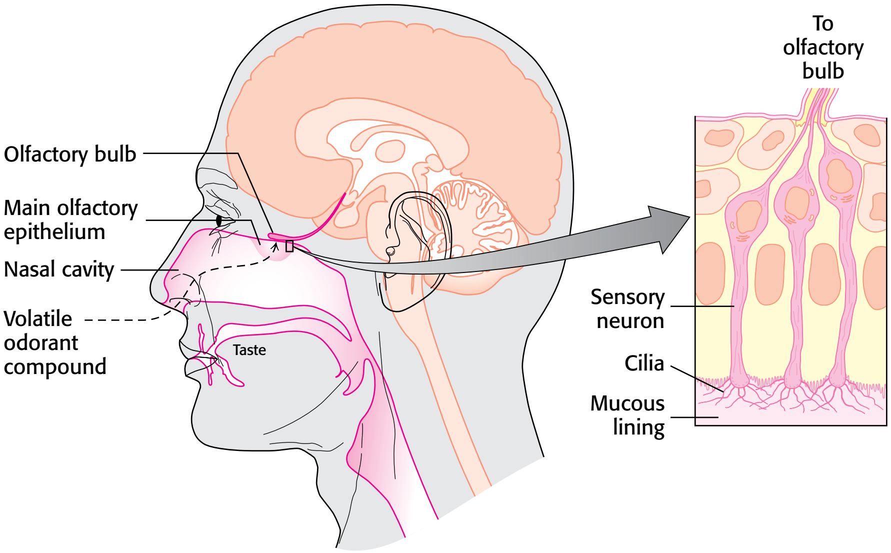


FIGURE 33.2 The main nasal epithelium. This region of the nose, which lies at the top of the nasal cavity, contains approximately 1 million sensory neurons. Nerve impulses generated by odorant molecules binding to receptors on the cilia travel from the sensory neurons to the olfactory bulb.

Biochemical studies in the late 1980s examined isolated cilia from rat olfactory epithelia that had been treated with odorants. Exposure to the odorants increased the cellular level of cyclic AMP, and this increase was observed only in the presence of GTP. On the basis of what was known about signal-transduction systems, *the participation of cAMP and GTP strongly suggested the involvement of a G protein and, hence, 7TM receptors.* Indeed, Randall Reed purified and cloned a G-protein α subunit, termed $G_{(olf)}$, which is uniquely expressed in olfactory cilia. The involvement of 7TM receptors suggested a strategy for identifying the olfactory receptors themselves. Complementary DNAs were sought that (1) were expressed primarily in the sensory neurons lining the nasal epithelium, (2) encoded members of the 7TM-receptor family, and (3) were present as a large and diverse family to account for the range of odorants. Through the use of these criteria, cDNAs for odorant receptors from rats were identified in 1991 by Richard Axel and Linda Buck.

The *odorant receptor* (hereafter, OR) family is even larger than expected: *more than 1000 OR genes are present in the mouse and the rat, whereas the human genome encodes approximately 350 ORs*. In addition, the human genome includes approximately 500 OR pseudogenes containing mutations that prevent the generation of a full-length, proper odorant receptor. The OR family is thus one of the largest gene families in human beings. Further analysis of primate OR genes reveals that the fraction of pseudogenes is greater in species more closely related to human beings (Figure 33.3). Thus, we may have a glimpse at the evolutionary loss of acuity in the sense of smell as higher mammals presumably became less dependent on this sense for survival. For rodents that are highly dependent on their sense of smell, essentially all OR genes encode functional proteins.

The OR proteins are typically 20% identical in sequence with the β-adrenergic receptor (Section 14.1) and from 30% to 60% identical with one another. Several specific sequence features are present in most or all OR family members (Figure 33.4). The central region,

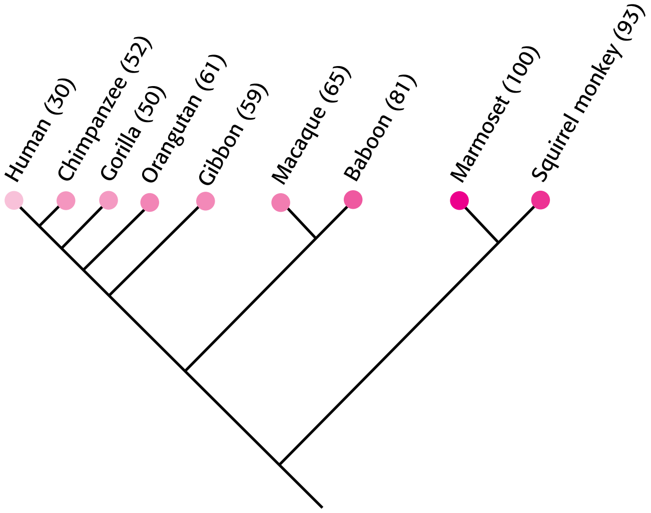


FIGURE 33.3 Evolution of odorant receptors. Odorant receptors appear to have lost function through conversion into pseudogenes in the course of primate evolution. The percentage of OR genes that appear to be functional for each species is given in parentheses.

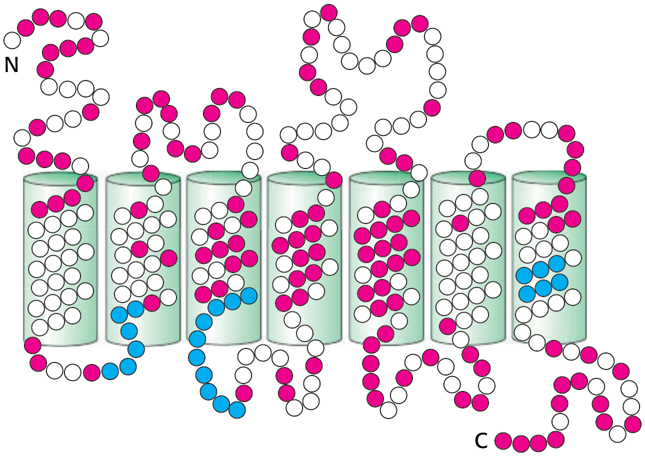


FIGURE 33.4 Conserved and variant regions in odorant receptors. Odorant receptors are members of the 7TM-receptor family. The green cylinders represent the seven presumed transmembrane helices. Strongly conserved residues characteristic of this protein family are shown in blue, whereas highly variable residues are shown in red.

particularly transmembrane helices 4 and 5, is highly variable, suggesting that this region is the site of odorant binding. That site must be different in odorant receptors that bind distinct odorant molecules.

What is the relation between OR gene expression and the individual neuron? Interestingly, *each olfactory neuron expresses only a single OR gene,* among hundreds available. Apparently, the precise OR gene expressed is determined largely at random. After one OR gene is expressed and a functional OR protein is produced, the expression of all other OR genes is suppressed by a feedback mechanism that remains to be fully elucidated.

The binding of an odorant to an OR on the neuronal surface initiates a signal-transduction cascade that results in an action potential (Figure 33.5). The ligand-bound OR activates $G_{(olf)}$, the specific G protein mentioned earlier. $G_{(olf)}$ is initially in its GDP-bound form. When activated, it releases GDP, binds GTP, and releases its associated $\beta\gamma$ subunits. The α subunit then activates a specific adenylate cyclase, increasing the intracellular concentration of cAMP. The rise in the intracellular concentration of cAMP activates a nonspecific cation channel that allows calcium and other cations into the cell. The flow of cations through the channel depolarizes the neuronal membrane and initiates an action potential. This action potential, combined with those from other olfactory neurons, leads to the perception of a specific odor.

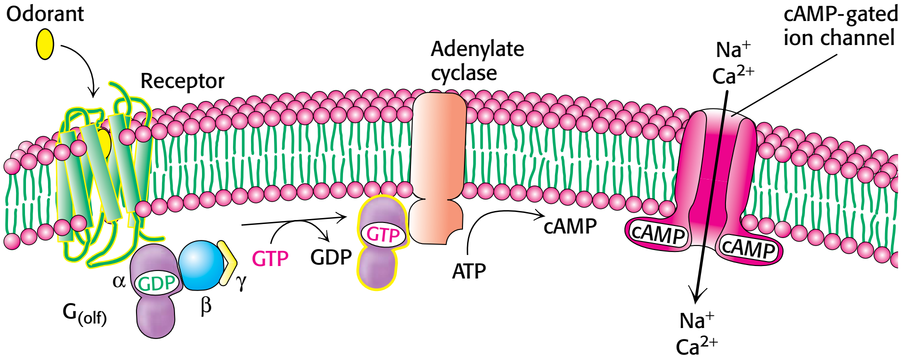


FIGURE 33.5 The olfactory signal-transduction cascade. The binding of odorant to the olfactory receptor activates a signaling pathway similar to those initiated in response to the binding of some hormones to their receptors. The final result is the opening of cAMP-gated ion channels and the initiation of an action potential.

Odorants are decoded by a combinatorial mechanism

An obvious challenge presented to an investigator by the large size of the OR family is to match each OR with the one or more odorant molecules to which it binds. Exciting progress has been made in this regard. Initially, an OR was matched with odorants by overexpressing a single, specific OR gene in rats. This OR responded to straight-chain aldehydes, most favorably to *n*-octanal and less strongly to *n*-heptanal and *n*-hexanal. More-dramatic progress was made by taking advantage of our knowledge of the OR signal-transduction pathway and the power of the polymerase chain reaction (PCR). A section of nasal epithelium from a mouse was loaded with the calcium-sensitive dye Fura-2 (Section 14.1). The tissue was then treated with different odorants, one at a time, at a specific concentration. If the

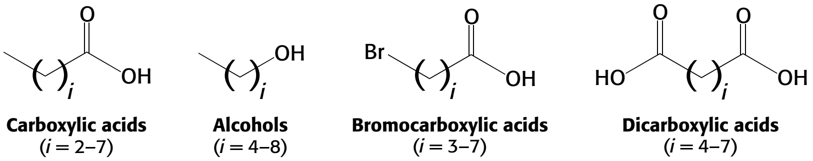
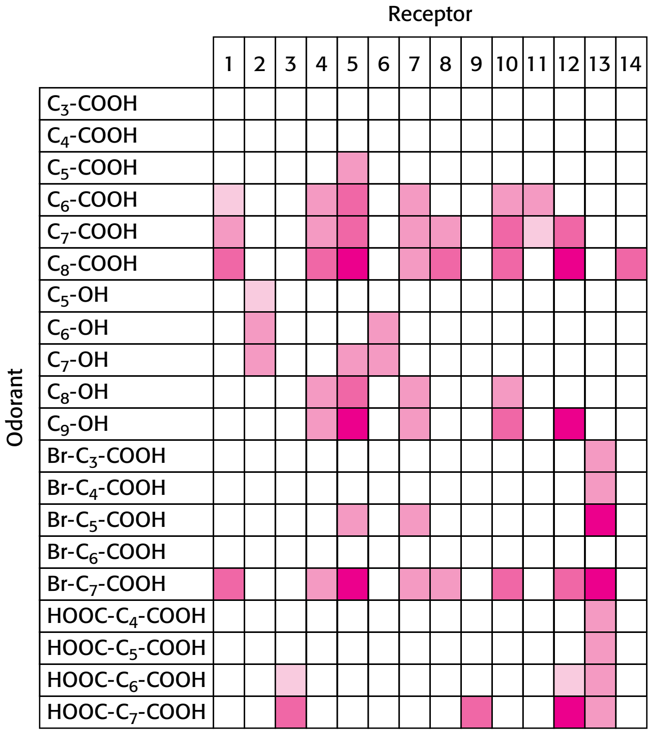


FIGURE 33.6 Four series of odorants tested for olfactory-receptor activation.

odorant had bound to an OR and activated it, that neuron could be detected under a microscope by the change in fluorescence caused by the influx of calcium that takes place as part of the signal-transduction process. To determine which OR was responsible for the response, cDNA was generated from mRNA that had been isolated from single identified neurons. The cDNA was then subjected to PCR with the use of primers that are effective in amplifying most or all OR genes. The sequence of the PCR product from each neuron was then determined and analyzed.

Using this approach, investigators analyzed the responses of neurons to a series of compounds having varying chain lengths and terminal functional groups (Figure 33.6). The results of these experiments appear surprising at first glance (Figure 33.7). Importantly, there is not a simple 1:1 correspondence between odorants and receptors. *Almost every odorant activates a number of receptors* (usually to different extents) and *almost every receptor is activated by more than one odorant*. Note, however, that each odorant activates a unique combination of receptors. In principle, this combinatorial mechanism allows even a small array of receptors to distinguish a vast number of odorants.



Odorant \ Receptor	1	2	3	4	5	6	7	8	9	10	11	12	13	14
C_3-COOH														
C_4-COOH														
C_5-COOH					▒									
C_6-COOH	░			▒	▓		▒			▒	▒			
C_7-COOH	▒			▒	▓		▒	▒		▓	░	▓		
C_8-COOH	▓			▓	█		▒	▓		▓		█		▓
C_5-OH		░												
C_6-OH		▒				▒								
C_7-OH		▒			▒	▒								
C_8-OH				▒	▓		▒			▒				
C_9-OH				▒	█		▒			▓		█		
Br-C_3-COOH													▒	
Br-C_4-COOH													▒	
Br-C_5-COOH					▒		▒						█	
Br-C_6-COOH														
Br-C_7-COOH	▓			▒	█		▒	▒		▓		▓	█	
HOOC-C_4-COOH													▒	
HOOC-C_5-COOH													▒	
HOOC-C_6-COOH			░									░	▒	
HOOC-C_7-COOH			▓						▓			█	▒	

FIGURE 33.7 Patterns of olfactory-receptor activation. Fourteen different receptors were tested for responsiveness to the compounds shown in Figure 33.6. A colored box indicates that the receptor at the top responded to the compound at the left. Darker colors indicate that the receptor was activated at a lower concentration of odorant.

How is the information about which receptors have been activated transmitted to the brain? Recall that each neuron expresses only one OR and that the pattern of expression appears to be largely random. A substantial clue to the connections between receptors and the brain has been provided by the creation of mice that express a gene for an easily detectable colored marker in conjunction with a specific OR gene. Olfactory neurons that express the OR–marker-protein combination were traced to their destination in the brain, a structure called the olfactory bulb (Figure 33.8). The processes from neurons that express the same OR gene were found to connect to the same location in the olfactory bulb. Moreover, this pattern of neuronal connection was found to be identical in all mice examined. Thus, *neurons that express specific ORs are linked to specific sites in the brain*. This property creates a spatial map of odorant-responsive neuronal activity within the olfactory bulb.

Can such a combinatorial mechanism truly distinguish many different odorants? An electronic "nose" that functions by the same principles

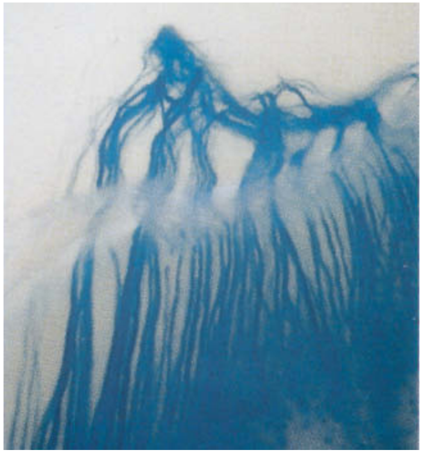


FIGURE 33.8 Converging olfactory neurons. This section of the nasal cavity is stained to reveal processes from sensory neurons expressing the same olfactory receptor. The processes converge to a single location in the olfactory bulb. [Reprinted from *Cell*, 87, P. Mombaerts, F. Wang, C. Dulac, S. K. Chao, A. Nemes, M. Mendelsohn, J. Edmondson, & R. Axel, Visualizing an Olfactory Sensory Map, p. 675–689, Copyright 1996, with permission from Elsevier.]

provides compelling evidence that it can (Figure 33.9). The receptors for the electronic nose are polymers that bind a range of small molecules. Each polymer binds every odorant, but to varying degrees. Importantly, the electrical properties of these polymers change on odorant binding. A set of 32 of these polymer sensors, wired together so that the pattern of responses can be evaluated, is capable of distinguishing individual compounds such as *n*-pentane and *n*-hexane as well as complex mixtures such as the odors of fresh and spoiled fruit.

FIGURE 33.9 The Cyranose 320. The electronic nose may find uses in the food industry, animal husbandry, law enforcement, and medicine. [Courtesy of Smiths Medical.]

33.2 Taste Is a Combination of Senses That Function by Different Mechanisms

The inability to taste food is a common complaint when nasal congestion reduces the sense of smell. Thus, smell greatly augments our sense of taste (also known as *gustation*), and taste is, in many ways, the sister sense to olfaction. Nevertheless, the two senses differ from each other in several important ways. First, we are able to sense several classes of compounds by taste that we are unable to detect by smell; salt and sugar have very little odor, yet they are primary stimuli of the gustatory system. Second, whereas we are able to discriminate thousands of odorants, discrimination by taste is much more modest. Five primary tastes are perceived: *bitter, sweet, sour, salty,* and *umami* (the taste of glutamate and aspartate from the Japanese word for "deliciousness"). These five tastes serve to classify compounds into potentially nutritive and beneficial (sweet, salty, umami) or potentially harmful or toxic (bitter, sour). Tastants (the molecules sensed by taste) are quite distinct for the different groups (Figure 33.10).

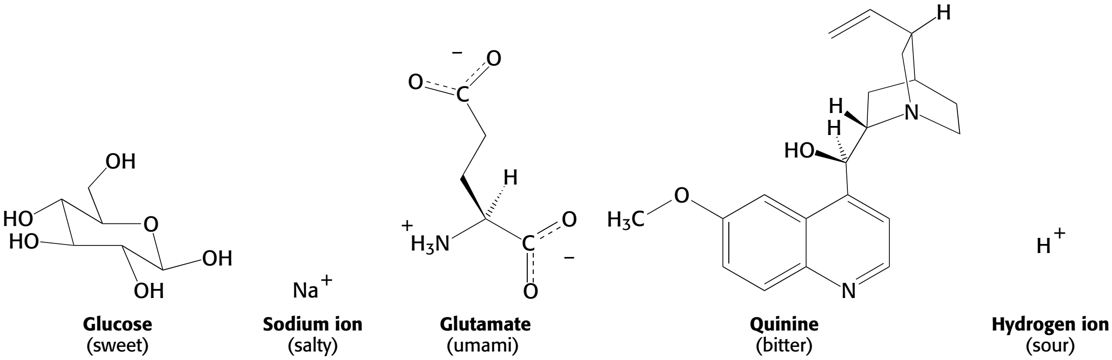


FIGURE 33.10 Examples of tastant molecules. Tastants fall into five groups: sweet, salty, umami, bitter, and sour.

The simplest tastant, the hydrogen ion, is perceived as sour. Other simple ions, particularly sodium ion, are perceived as salty. The taste called umami is evoked by the amino acids glutamate and aspartate, the former often encountered as the flavor enhancer monosodium glutamate (MSG). In contrast, *tastants perceived as sweet and, particularly, bitter are extremely diverse*. Many bitter compounds are alkaloids or other plant products, many of which are toxic. However, they do not have any common structural elements or other common properties. Carbohydrates such as glucose and sucrose are perceived as sweet, as are other compounds including some simple peptide derivatives, such as aspartame, and even some proteins.

These differences in specificity among the five tastes are due to differences in their underlying biochemical mechanisms. The sense of taste is, in fact, a number of independent senses all utilizing the same organ, the tongue, for their expression.

Tastants are detected by specialized structures called *taste buds,* which contain approximately 150 cells, including sensory neurons (Figure 33.11). Fingerlike projections called *microvilli,* which are rich in taste receptors, project from one end of each sensory neuron to the surface of the tongue. Nerve fibers at the opposite end of each neuron carry electrical impulses to the brain in response to stimulation by tastants. Structures called *taste papillae* contain numerous taste buds.

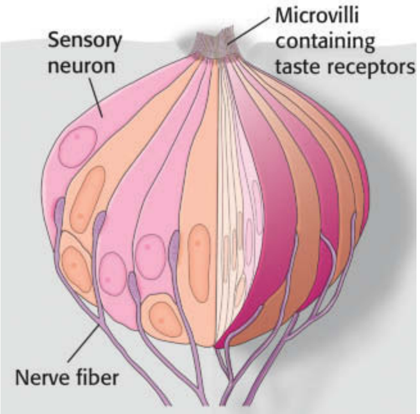
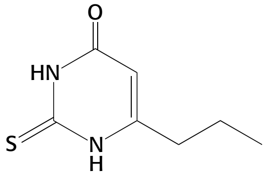


FIGURE 33.11 A taste bud. Each taste bud contains sensory neurons that extend microvilli to the surface of the tongue, where they interact with tastants.

Sequencing of the human genome led to the discovery of a large family of 7TM bitter receptors

Just as in olfaction, a number of clues pointed to the involvement of G proteins and, hence, 7TM receptors in the detection of bitter and sweet tastes. The evidence included the isolation of a specific G-protein α subunit termed *gustducin,* which is expressed primarily in taste buds (Figure 33.12). How could the 7TM receptors be identified? The ability to detect some compounds depends on specific genetic loci in both human beings and mice. For instance, the ability to taste the bitter compound 6-*n*-propyl-2-thiouracil (PROP) was mapped to a region on human chromosome 5 by comparing DNA markers of persons who vary in sensitivity to this compound.



6-*n*-Propyl-2-thiouracil (PROP)

This observation suggested that this region might encode a 7TM receptor that responded to PROP. Approximately 450 kilobases in this region had been sequenced early in the human genome project. This sequence was searched by computer for potential 7TM-receptor genes, and, indeed, one was detected and named T2R1. Additional database searches detected approximately 30 sequences similar to T2R1 in the human genome. The encoded proteins are between 30 and 70% identical with T2R1 (Figure 33.13).

Are these proteins, in fact, bitter receptors? Several lines of evidence suggest that they are. First, their genes are expressed in taste-sensitive cells—in fact, in many of the same cells that express gustducin. Second, cells that express individual members of this family respond to specific bitter compounds. For example, cells that express a specific mouse receptor (mT2R5) responded when exposed specifically to cycloheximide. Third, mice that had been found unresponsive to cycloheximide were found to have point mutations in the gene encoding mT2R5. Finally, cycloheximide

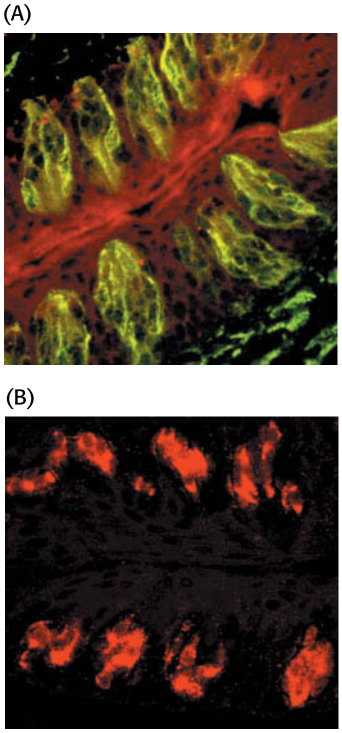


FIGURE 33.12 Expression of gustducin in the tongue. (A) A section of tongue stained with a fluorescent antibody reveals the position of the taste buds. (B) The same region stained with an antibody directed against gustducin reveals that this G protein is expressed in taste buds. [Courtesy of Dr. Charles S. Zuker.]

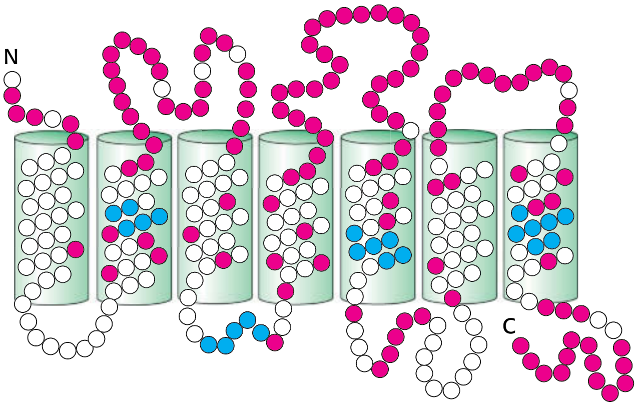


FIGURE 33.13 Conserved and variant regions in bitter receptors. The bitter receptors are members of the 7TM-receptor family. Strongly conserved residues characteristic of this protein family are shown in blue, and highly variable residues are shown in red.

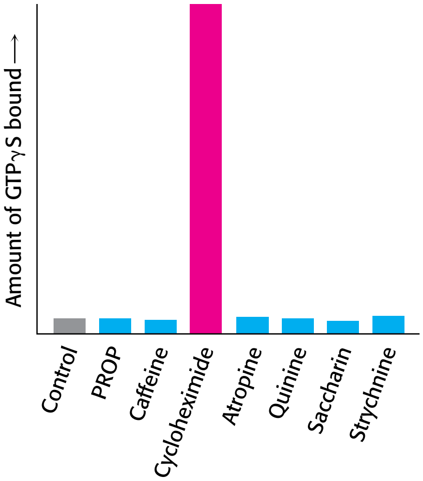


FIGURE 33.14 Evidence that T2R proteins are bitter taste receptors. Cycloheximide uniquely stimulates the binding of the GTP analog GTPγS to gustducin in the presence of the mT2R protein. [Information from J. Chandrashekar et al., *Cell* 100:703–711, 2000.]

specifically stimulates the binding of GTP analogs to gustducin in the presence of the mT2R5 protein (Figure 33.14).

Importantly, each taste-receptor cell expresses many different members of the T2R family. This pattern of expression stands in sharp contrast to the pattern of one receptor type per cell that characterizes the olfactory system (Figure 33.15). The difference in expression patterns accounts for the much greater specificity of our perceptions of odors compared with tastes. *We are able to distinguish among subtly different odors because each odorant stimulates a unique pattern of neurons. In contrast, many tastants stimulate the same neurons*. Thus, we perceive only "bitter" without the ability to discriminate cycloheximide from quinine.

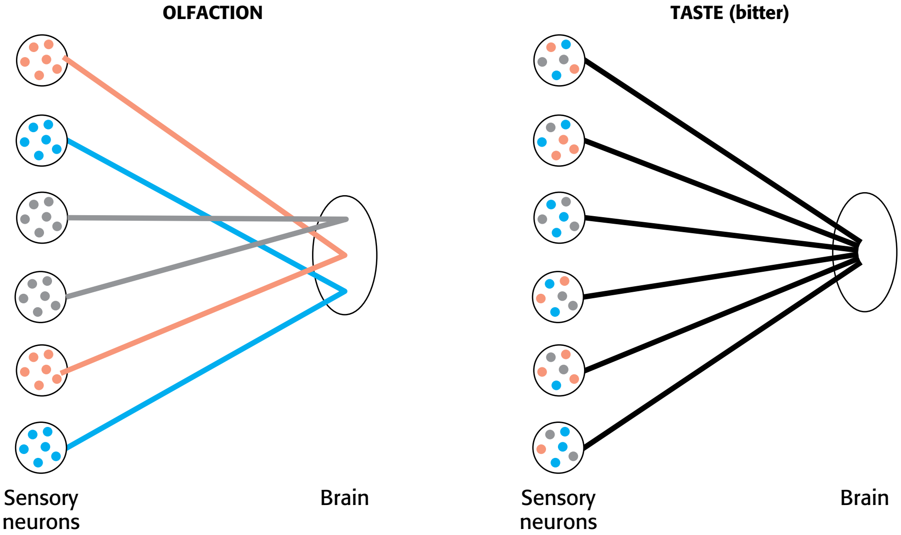


FIGURE 33.15 Differing gene-expression and connection patterns in olfactory and bitter taste receptors. In olfaction, each neuron expresses a single OR gene, and the neurons expressing the same OR converge to specific sites in the brain, enabling specific perception of different odorants. In gustation, each neuron expresses many bitter receptor genes, and so the identity of the tastant is lost in transmission.

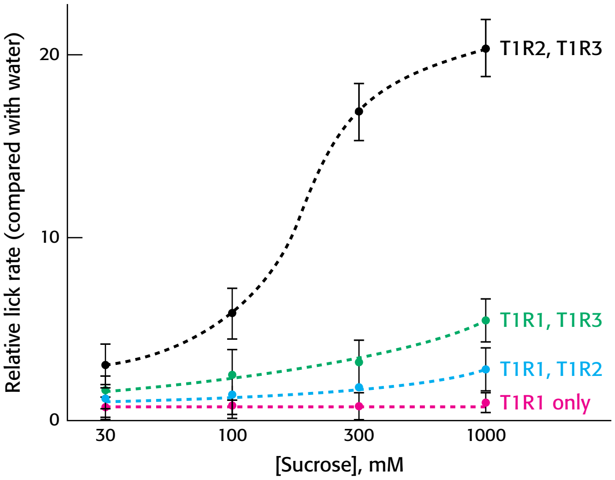


FIGURE 33.16 Evidence for a heterodimeric sweet receptor. The sensitivity to sweetness of mice with genes for T1R1 only, both T1R1 and T1R2, both T1R1 and T1R3, or both T1R2 and T1R3 was determined by observing the relative rates at which they licked solutions containing various amounts of sucrose. These studies revealed that both T1R2 and T1R3 were required for a full response to sucrose. [Data from G. Q. Zhao et al., *Cell* 115:255–266, 2003.]

A heterodimeric 7TM receptor responds to sweet compounds

Most sweet compounds are carbohydrates, energy rich and easily digestible. Some noncarbohydrate compounds such as saccharin and aspartame also taste sweet. Members of a second family of 7TM receptors are expressed in taste-receptor cells sensitive to sweetness. The three members of this family, referred to as T1R1, T1R2, and T1R3, are distinguished by their large extracellular domains compared with those of the bitter receptors. Studies in knockout mice have revealed that T1R2 and T1R3 are expressed simultaneously in mice able to taste carbohydrates (Figure 33.16). Thus, T1R2 and T1R3 appear to form a specific heterodimeric receptor responsible for mediating the response to sugars. This heterodimeric receptor also responds to artificial sweeteners and to sweet-tasting proteins and therefore appears to be the receptor responsible for responses to all sweet tastants. Note that T1R2 and T1R3 do respond to sweet tastants individually but only at very high concentrations of tastant.

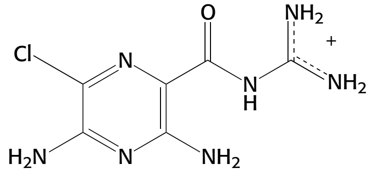
The requirement for an *oligomeric* 7TM receptor for a fully functional response is surprising, considering our previous understanding of 7TM receptors. This discovery has at least two possible explanations. First, the sweet receptor could be a member of a small subset of the 7TM-receptor family that functions well only as oligomers. Alternatively, many 7TM receptors may function as oligomers, but this notion is not clear, because these oligomers contain only one type of 7TM-receptor subunit. Further studies will be required to determine which of these explanations is correct.

Umami, the taste of glutamate and aspartate, is mediated by a heterodimeric receptor related to the sweet receptor

The family of receptors responsible for detecting sweetness is also responsible for detecting amino acids. In human beings, only glutamate and aspartate elicit a taste response. Studies similar to those for the sweet receptor revealed that the umami receptor consists of T1R1 and T1R3. Thus, this receptor has one subunit (T1R3) in common with the sweet receptor but has an additional subunit (T1R1) that does not participate in the sweet response. This conclusion is supported by the observation that mice in which the gene for T1R1 is disrupted do not respond to aspartate but do respond normally to sweet tastants; mice having disrupted genes for both T1R1 and T1R3 respond poorly to both umami and sweet tastants.

Salty tastes are detected primarily by the passage of sodium ions through channels

Salty tastants are not detected by 7TM receptors. Rather, they are detected directly by their passage through ion channels expressed on the surface of cells in the tongue. Evidence for the role of these ion channels comes from examining known properties of Na^+ channels characterized in other biological contexts. One class of channels, characterized first for its role in salt reabsorption, is thought to be important in the detection of salty tastes because these channels are sensitive to the compound *amiloride,* which mutes the taste of salt and significantly lowers sensory-neuron activation in response to sodium.



Amiloride

An *amiloride-sensitive* Na^+ *channel* comprises four subunits that may be either identical or distinct but in any case are homologous. An individual subunit ranges in length from 500 to 1000 amino acids and includes two presumed membrane-spanning helices as well as a large extracellular domain in between them (Figure 33.17). The extracellular region includes two (or, sometimes, three) distinct regions rich in cysteine residues (and, presumably, disulfide bonds). A region just ahead of the second membrane-spanning helix appears to form part of the pore in a manner analogous to that of the structurally characterized potassium channel. The members of the amiloride-sensitive Na^+ -channel family are numerous and diverse in their biological roles. We shall encounter them again in the context of the sense of touch.

Sodium ions passing through these channels produce a significant transmembrane current. Amiloride blocks this current, accounting for its effect on taste. However, about 20% of the response to sodium remains even in the presence of amiloride, suggesting that other ion channels also contribute to salt detection.

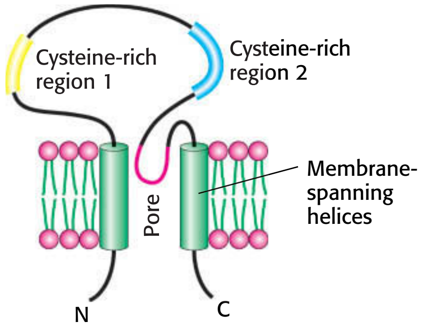


FIGURE 33.17 Schematic structure of the amiloride-sensitive sodium channel. Only one of the four subunits that constitute the functional channel is illustrated. The amiloride-sensitive sodium channel belongs to a superfamily having common structural features, including two hydrophobic membrane-spanning regions, intracellular amino and carboxyl termini, and a large, extracellular region with conserved cysteine-rich domains.

Sour tastes arise from the effects of hydrogen ions (acids) on channels

Like salty tastes, *sour tastes are detected by direct interactions with ion channels,* but the incoming ions are hydrogen ions (in high concentrations) rather than sodium ions. For example, in the absence of high concentrations of sodium, hydrogen ion flow can induce substantial transmembrane currents through amiloride-sensitive Na^+ channels. However, hydrogen ions are also sensed by mechanisms other than their direct passage through membranes. Binding by hydrogen ions blocks some potassium ion channels and activates other types of channels. Together, these mechanisms lead to changes in membrane polarization in sensory neurons that produce the sensation of sour taste. We shall consider an additional receptor related to taste, one responsible for the "hot" taste of spicy food, when we examine mechanisms of touch perception.

33.3 Photoreceptor Molecules in the Eye Detect Visible Light

Vision is based on the absorption of light by photoreceptor cells in the eye. These cells are sensitive to light in a narrow region of the electromagnetic spectrum, the region with wavelengths between 390 and 750 nm (Figure 33.18). Vertebrates have two kinds of photoreceptor cells, called *rods* and *cones* because of their distinctive shapes. Cones function in bright light and are responsible for color vision, whereas rods function in dim light but do not perceive color. A human retina contains about 3 million cones and 100 million rods. Remarkably, a rod cell can respond to a single photon, and the brain requires fewer than 10 such responses to register the sensation of a flash of light.

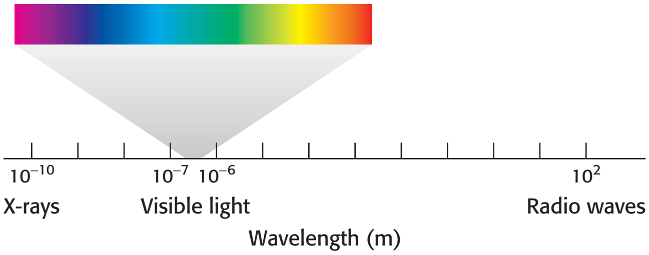
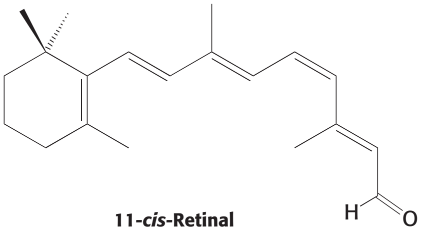


FIGURE 33.18 The electromagnetic spectrum. Visible light has wavelengths between 390 and 750 nm.

Rhodopsin, a specialized 7TM receptor, absorbs visible light



11-*cis*-Retinal

Rods are slender, elongated structures; the outer segment is specialized for photoreception (Figure 33.19). It contains a stack of about 1000 discs, which are membrane-enclosed sacs densely packed with photoreceptor molecules. The photosensitive molecule is often called a *visual pigment* because it is highly colored owing to its ability to absorb light. The photoreceptor molecule in rods is *rhodopsin* (Section 14.1), which consists of the protein *opsin* linked to *11-*cis*-retinal,* a prosthetic group.

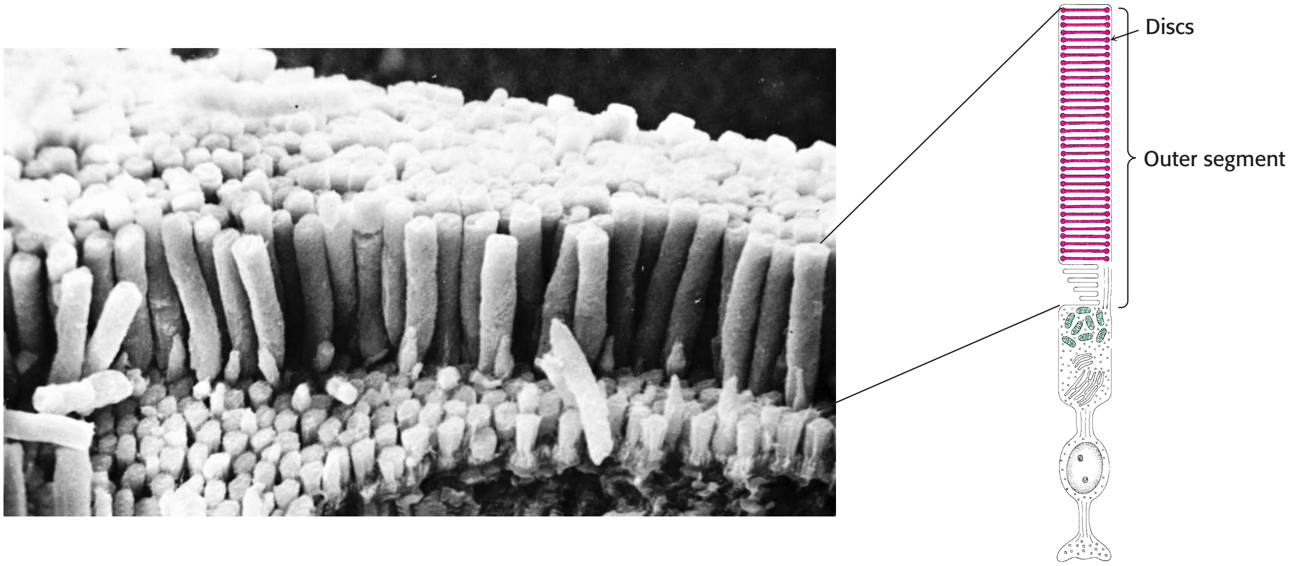


FIGURE 33.19 The rod cell. (Left) Scanning electron micrograph of retinal rod cells. (Right) Schematic representation of a rod cell. [Photograph courtesy of Dr. Deric Bownds.]

Rhodopsin absorbs light very efficiently in the middle of the visible spectrum, its absorption being centered on 500 nm, which nicely matches the solar output (Figure 33.20). A rhodopsin molecule will absorb a high percentage of the photons of the correct wavelength that strike it, as indicated by the extinction coefficient of 40,000 $M^{-1}cm^{-1}$ at 500 nm. The extinction coefficient for rhodopsin is more than an order of magnitude greater than that for tryptophan, the most efficient absorber in proteins that lack prosthetic groups.

Opsin, the protein component of rhodopsin, is a member of the 7TM-receptor family. Indeed, rhodopsin was the first member of this family to be purified, its gene was the first to be cloned and sequenced, and its three-dimensional structure was the first to be determined. The color of rhodopsin and its responsiveness to light depend on the presence of the light-absorbing group *(chromophore)* 11-*cis*-retinal. This compound is a powerful absorber of light because it is a polyene; its six alternating single and double bonds constitute a long, unsaturated electron network. Recall that alternating single and double bonds account for the chromophoric properties of chlorophyll (Section 19.2). The aldehyde group of 11-*cis*-retinal forms a Schiff base (Figure 33.21) with the ε-amino group of lysine residue 296, which lies in the center of the seventh transmembrane helix. Free retinal absorbs maximally at 370 nm, and its unprotonated Schiff-base adduct absorbs at 380 nm, whereas the protonated Schiff base absorbs at 440 nm or longer wavelengths. Thus, *the 500-nm absorption maximum for rhodopsin strongly suggests that the Schiff base is protonated;* additional interactions with opsin shift the absorption maximum farther toward the red. The positive charge of the protonated Schiff base is compensated by the negative charge of glutamate 113 located in helix 2; the glutamate residue closely approaches the lysine–retinal linkage in the three-dimensional structure of rhodopsin.

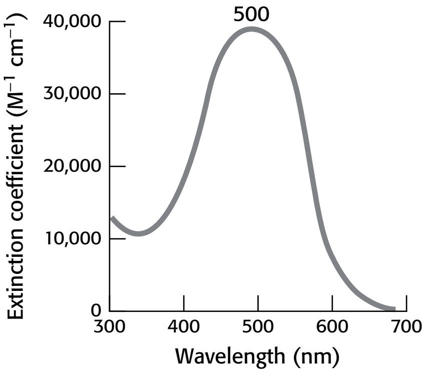
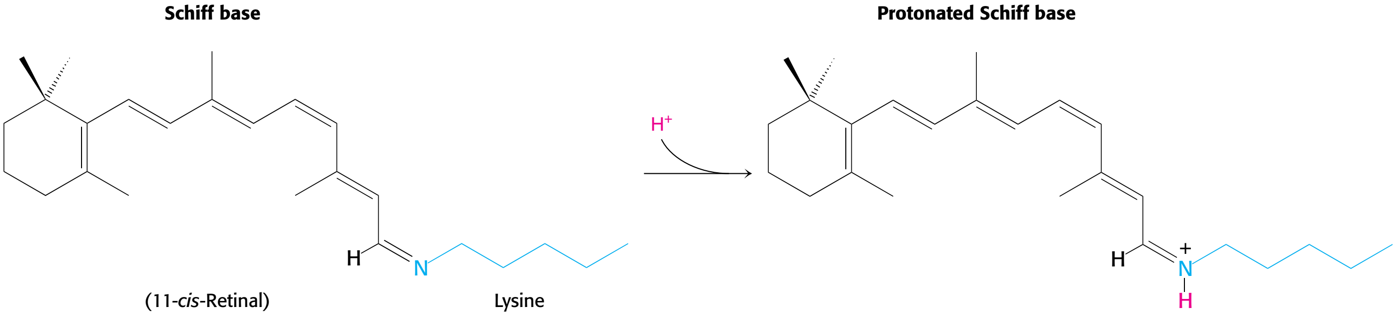


FIGURE 33.20 Rhodopsin absorption spectrum. Almost all photons with wavelengths near 500 nm that strike a rhodopsin molecule are absorbed.

FIGURE 33.21 Retinal–lysine linkage. Retinal is linked to lysine 296 in opsin by a Schiff-base linkage. In the resting state of rhodopsin, this Schiff base is protonated.



Light absorption induces a specific isomerization of bound 11-*cis*-retinal

How does the absorption of light by the retinal Schiff base generate a signal? George Wald and his coworkers discovered that *light absorption results in the isomerization of the 11-*cis*-retinal group of rhodopsin to its all-*trans* form* (Figure 33.22). This isomerization causes the Schiff-base nitrogen atom to move approximately 5 Å, assuming that the cyclohexane ring of the retinal group remains fixed. In essence, *the light energy of a photon is converted into atomic motion*. The change in atomic positions, like the binding of a ligand to other 7TM receptors, sets in train a series of events that lead to the closing of ion channels and the generation of a nerve impulse.

The isomerization of the retinal Schiff base takes place within a few picoseconds of a photon being absorbed. The initial product, termed *bathorhodopsin,* contains a strained all-*trans*-retinal group. Within approximately

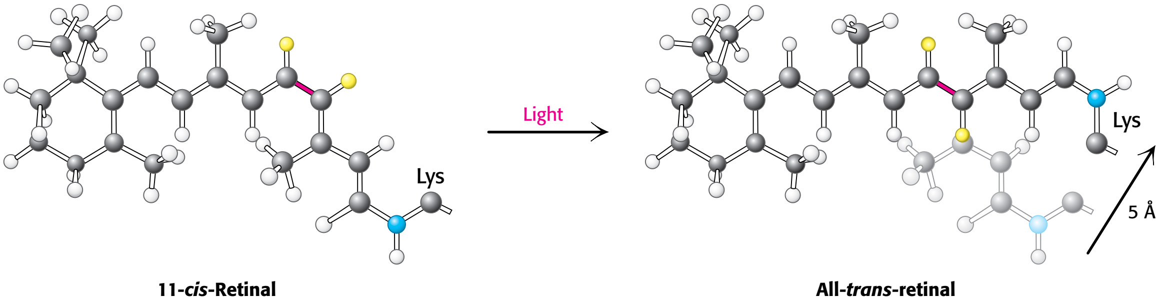


FIGURE 33.22 Atomic motion in retinal. The Schiff-base nitrogen atom moves 5 Å as a consequence of the light-induced isomerization of 11-*cis*-retinal to all-*trans*-retinal by rotation about the bond shown in red.

1 ms, this intermediate is converted through several additional intermediates into *metarhodopsin II*. In metarhodopsin II, the Schiff base is deprotonated and the opsin protein has undergone significant reorganization.

Metarhodopsin II (also referred to as R*) is analogous to the ligand-bound state of 7TM receptors such as the β_2-adrenergic receptor (Section 14.1) and the odorant and tastant receptors discussed previously (Figure 33.23). Like these receptors, this form of rhodopsin activates a heterotrimeric G protein that propagates the signal. The G protein associated with rhodopsin is called *transducin*. Metarhodopsin II triggers the exchange of GDP for GTP by the α subunit of transducin (Figure 33.24). On the binding of GTP, the βγ subunits of transducin are released and the α subunit switches on a *cGMP phosphodiesterase* by binding to an inhibitory subunit and removing it. The activated phosphodiesterase is a potent enzyme that rapidly hydrolyzes cGMP to GMP. The reduction in cGMP concentration causes *cGMP-gated ion channels* to close, leading to the hyperpolarization of the membrane and neuronal signaling. *At each step in this process, the initial signal—the absorption of a single photon—is amplified so that it leads to sufficient membrane hyperpolarization to result in signaling.*

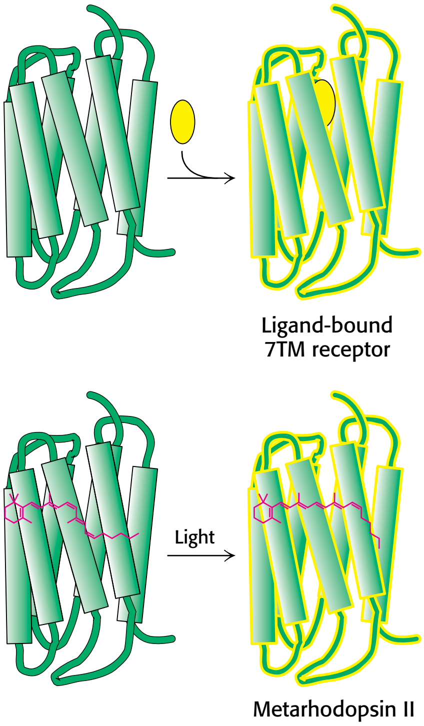


FIGURE 33.23 Analogous 7TM receptors. The conversion of rhodopsin into metarhodopsin II activates a signal-transduction pathway analogously to the activation induced by the binding of other 7TM receptors to appropriate ligands.

Light-induced lowering of the calcium level coordinates recovery

As we have seen, the visual system responds to changes in light and color within a few milliseconds, quickly enough that we are able to perceive continuous motion at nearly 1000 frames per second. To achieve a rapid response, the signal must also be terminated rapidly and the system must be returned to its initial state. First, activated rhodopsin must be blocked from continuing to activate transducin. *Rhodopsin kinase* catalyzes the phosphorylation of the carboxyl terminus of R* at multiple serine and threonine residues. *Arrestin,* an inhibitory protein then binds phosphorylated R* and prevents additional interaction with transducin.

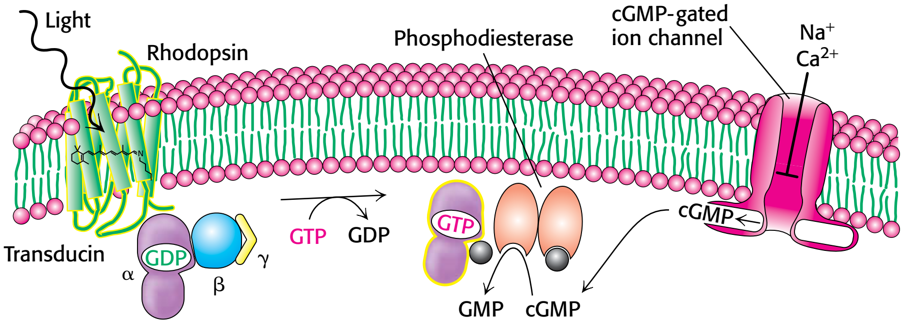


FIGURE 33.24 Visual signal transduction. The light-induced activation of rhodopsin leads to the hydrolysis of cGMP, which in turn leads to ion-channel closing and the initiation of an action potential.

Second, the α subunit of transducin must be returned to its inactive state to prevent further signaling. Like other G proteins, the α subunit possesses built-in GTPase activity that hydrolyzes bound GTP to GDP. Hydrolysis takes place in less than a second when transducin is bound to the phosphodiesterase. The GDP form of transducin then leaves the phosphodiesterase and reassociates with the βγ subunits, and the phosphodiesterase returns to its inactive state. Third, the level of cGMP must be raised to reopen the cGMP-gated ion channels. *The action of guanylate cyclase accomplishes this third step by synthesizing cGMP from GTP.*

Calcium ion plays an essential role in controlling guanylate cyclase because it markedly inhibits the activity of the enzyme. In the dark, Ca^{2+} as well as Na^{+} enter the rod outer segment through the cGMP-gated channels. Calcium ion influx is balanced by its efflux through an exchanger, a transport system that uses the thermodynamically favorable flow of four Na^{+} ions into the cell and one K^{+} ion out of the cell to extrude one Ca^{2+} ion. After illumination, the entry of Ca^{2+} through the cGMP-gated channels stops, but its export through the exchanger continues. Thus, the cytoplasmic Ca^{2+} level drops from 500 nM to 50 nM after illumination. This drop markedly stimulates guanylate cyclase, rapidly restoring the concentration of cGMP to reopen the cGMP-gated channels.

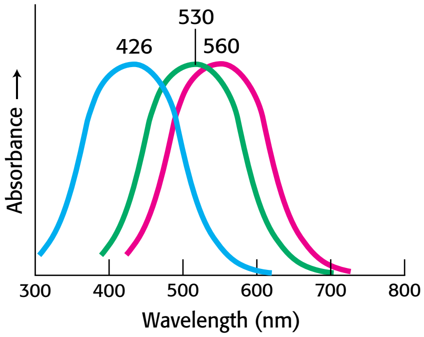


FIGURE 33.25 Cone-pigment absorption spectra. The absorption spectra of the cone visual pigment responsible for color vision.

$$\underbrace{[\text{cGMP}]\downarrow \longrightarrow \begin{matrix}\text{Ion}\\ \text{channels}\\ \text{closed}\end{matrix}}_{\text{Activation}} \longrightarrow \underbrace{[\text{Ca}^{2+}]\downarrow \longrightarrow \begin{matrix}\text{Guanylate}\\ \text{cyclase}\\ \text{activity}\\ \text{increased}\end{matrix} \longrightarrow [\text{cGMP}]\uparrow}_{\text{Recovery}}$$

By controlling the rate of cGMP synthesis, Ca^{2+} levels govern the speed with which the system is restored to its initial state.

Color vision is mediated by three cone receptors that are homologs of rhodopsin

Cone cells, like rod cells, contain visual pigments. Like rhodopsin, these photoreceptor proteins are members of the 7TM-receptor family and use 11-*cis*-retinal as their chromophore. In human cone cells, there are three distinct photoreceptor proteins with absorption maxima at 426, 530, and ~560 nm (Figure 33.25). *These absorbances correspond to the blue, green, and yellow-green regions of the spectrum*. Recall that the absorption maximum for rhodopsin is 500 nm. The photoreceptor protein with its absorption maximum near 560 nm absorbs red light (with wavelength >620 nm) whereas the other two photoreceptors do not. For simplicity, we shall refer to the three proteins as the blue, green, and red photoreceptors.

The amino acid sequences of the cone photoreceptors have been compared with one another and with rhodopsin. The result is striking. Each of the cone photoreceptors is approximately 40% identical in sequence with rhodopsin. Similarly, the blue photoreceptor is 40% identical with each of the green and red photoreceptors. The green and red photoreceptors, however, are >95% identical with each other, differing in only 15 of 364 positions (Figure 33.26).

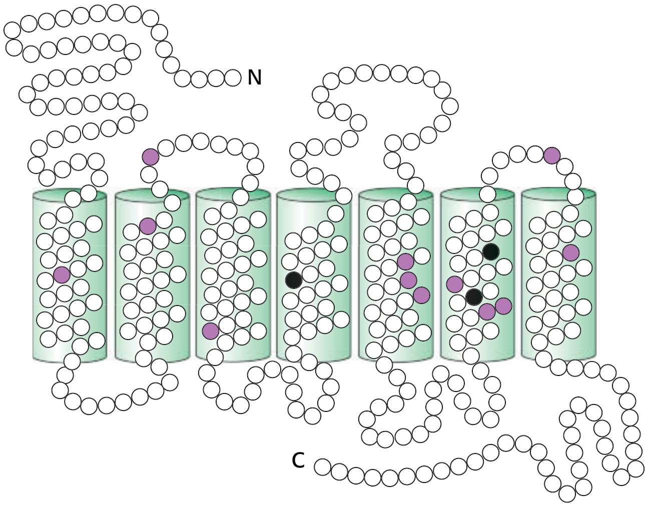


FIGURE 33.26 Comparison of the amino acid sequences of the green and red photoreceptors. Open circles correspond to identical residues, whereas colored circles mark residues that are different. The differences in the three black positions are responsible for most of the difference in their absorption spectra.

These observations are sources of insight into photoreceptor evolution. First, the green and red

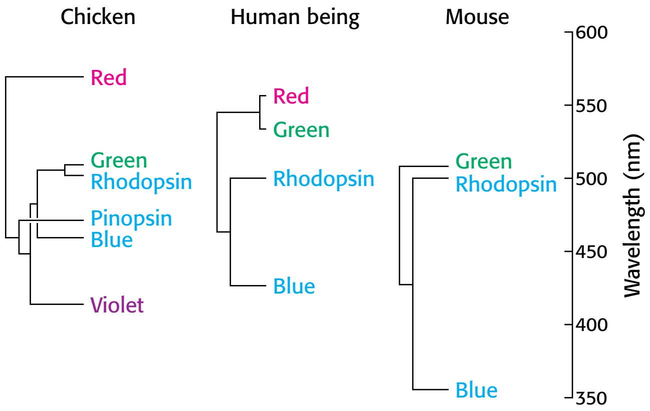


FIGURE 33.27 Evolutionary relationships among visual pigments. Visual pigments have evolved by gene duplication along different branches of the animal evolutionary tree. The branch lengths of the "trees" correspond to the percentage of amino acid divergence. [Information from J. Nathans, *Neuron* 24:299–312, 1999; by permission of Cell Press.]

photoreceptors are clearly products of a recent evolutionary event (Figure 33.27). The green and red pigments appear to have diverged in the primate lineage approximately 35 million years ago. Mammals, such as dogs and mice, that diverged from primates earlier have only two cone photoreceptors, blue and green. They are not sensitive to light as far toward the infrared region as we are, and they do not discriminate colors as well. In contrast, birds such as chickens have a total of six pigments: rhodopsin, four cone pigments, and a pineal visual pigment called *pinopsin*. Birds have highly acute color perception.

Second, the high level of similarity between the green and the red pigments has made the identification of the specific amino acid residues responsible for spectral tuning possible. Three residues (at positions 180, 277, and 285) are responsible for most of the difference between the green and the red pigments. In the green pigment, these residues are alanine, phenylalanine, and alanine, respectively; in the red pigment, they are serine, tyrosine, and threonine. A hydroxyl group has been added to each amino acid in the red pigment. The hydroxyl groups can interact with the photoexcited state of retinal and lower its energy, leading to a shift toward the lower-energy (red) region of the spectrum.

Homologous recombination
The exchange of DNA segments with substantial sequence similarity at equivalent positions between chromosomes.

Rearrangements in the genes for the green and red pigments lead to "color blindness"

The genes for the green and red pigments lie adjacent to each other on the human X chromosome. These genes are more than 98% identical in nucleotide sequence, including introns and untranslated regions as well as the protein-coding region. Regions with such high similarity are very susceptible to unequal homologous recombination.

Recombination can take place either between or within transcribed regions of the gene (Figure 33.28). If recombination takes place between transcribed regions, the product chromosomes will differ in the number of pigment genes that they carry. One chromosome will lose a gene and thus may lack the gene for, say, the green pigment; the other chromosome will gain a gene. Consistent with this scenario, approximately 2% of human X chromosomes carry only a single color-pigment gene, approximately 20% carry two, 50% carry three, 20% carry four, and 5% carry five or more. A person lacking the gene for the green pigment will have trouble distinguishing red and green color, characteristic of the most common form of color blindness. Approximately 5% of males have this form of color blindness. Recombination can also take place within the

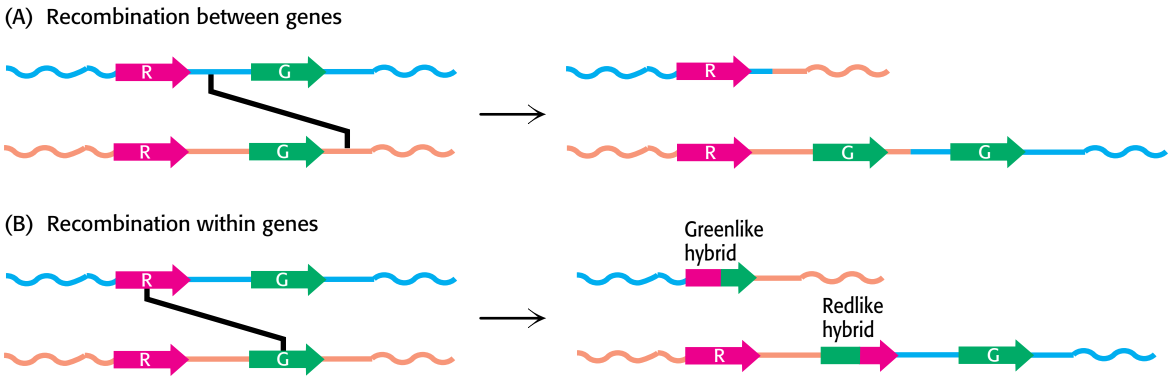

(A) Recombination between genes
R
G
R
G
R
R
G
G
(B) Recombination within genes
R
G
R
G
Greenlike hybrid
Redlike hybrid
R
G

FIGURE 33.28 Recombination pathways leading to color blindness. Rearrangements in the course of DNA replication may lead to (A) the loss of visual pigment genes or (B) the formation of hybrid pigment genes that encode photoreceptors with anomalous absorption spectra. Because the amino acids most important for determining absorption spectra are in the carboxyl-terminal half of each photoreceptor protein, the part of the gene that encodes this region most strongly affects the absorption characteristics of hybrid receptors. [Information from J. Nathans, *Neuron* 24:299–312, 1999; by permission of Cell Press.]

transcription units, resulting in genes that encode hybrids of the green and red photoreceptors. The absorption maximum of such a hybrid lies between that of the red and green pigments. A person with such hybrid genes who also lacks either a functional red- or green-pigment gene does not discriminate color well.

33.4 Hearing Depends on the Speedy Detection of Mechanical Stimuli

Hearing and touch are based on the detection of mechanical stimuli. Although the proteins of these senses have not been as well characterized as those of the senses already discussed, anatomical, physiological, and biophysical studies have elucidated the fundamental processes. *A major clue to the mechanism of hearing is its speed.* We hear frequencies ranging from 200 to 20,000 Hz (cycles per second), corresponding to times of 5 to 0.05 ms. Furthermore, our ability to locate sound sources, one of the most important functions of hearing, depends on the ability to detect the time delay between the arrival of a sound at one ear and its arrival at the other. Given the separation of our ears and the speed of sound, we must be able to accurately sense time differences of 0.7 ms. In fact, human beings can locate sound sources associated with temporal delays as short as 0.02 ms. This high time resolution implies that hearing must employ direct transduction mechanisms that do not depend on second messengers. Recall that, in vision, for which speed also is important, the signal-transduction processes take place in milliseconds.

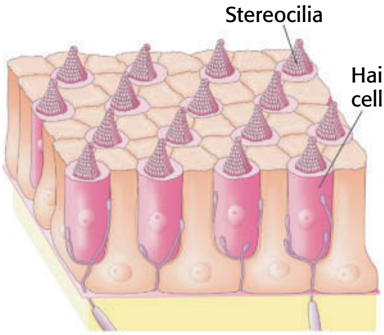


FIGURE 33.29 Hair cells, the sensory neurons crucial for hearing. These specialized neurons are capped with hairlike projections called stereocilia that are responsible for detecting very subtle vibrations. [Information from A. J. Hudspeth, *Nature* 341:397–404, 1989.]

Hair cells use a connected bundle of stereocilia to detect tiny motions

Sound waves are detected inside the cochlea of the inner ear. The *cochlea* is a fluid-filled, membranous sac that is coiled like a snail shell. The primary detection is accomplished by specialized neurons inside the cochlea called *hair cells* (Figure 33.29). Each cochlea contains approximately 16,000 hair cells, and each hair cell contains a hexagonally shaped bundle of 20 to 300 hairlike projections called *stereocilia* (Figure 33.30). These stereocilia are graded in length across the bundle. Mechanical deflection of the hair bundle, as takes place when a sound wave arrives at the ear, creates a change in the membrane potential of the hair cell.

Micromanipulation experiments have directly probed the connection between mechanical stimulation and membrane potential. Displacement toward the direction of the tallest part of the hair bundle results in the depolarization of the hair cell, whereas displacement in the opposite direction results in its hyperpolarization (Figure 33.31). Motion perpendicular to the hair-length gradient does not produce any change in resting potential.

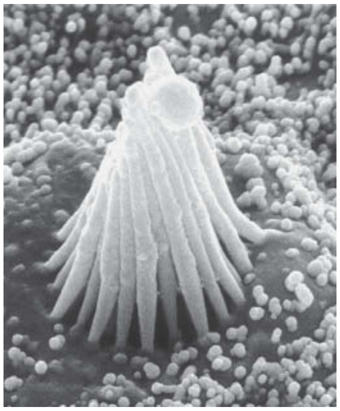


FIGURE 33.30 An electron micrograph of a hair bundle. [Courtesy of Dr. A. Jacobs and Dr. A. J. Hudspeth.]

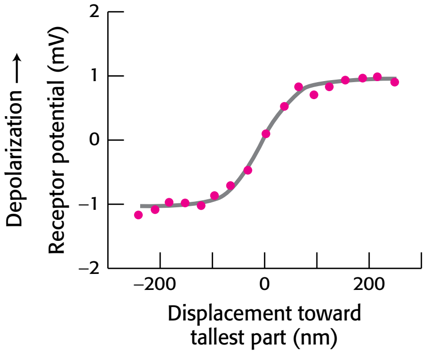


FIGURE 33.31 Micromanipulation of a hair cell. Movement toward the tallest part of the bundle depolarizes the cell as measured by the microelectrode. Movement toward the shortest part hyperpolarizes the cell. Lateral movement has no effect. [After A. J. Hudspeth, *Nature* 341:397–404, 1989.]

Remarkably, *displacement of the hair bundle by as little as 3 Å (0.3 nm) results in a measurable (and functionally important) change in membrane potential*. This motion of 0.001 degree corresponds to a 0.3 inch movement of the top of the Empire State Building.

How does the motion of the hair bundle create a change in membrane potential? The rapid response, within microseconds, suggests that the movement of the hair bundle acts on ion channels directly. An important observation is that adjacent stereocilia are linked by individual filaments called *tip links* (Figure 33.32).

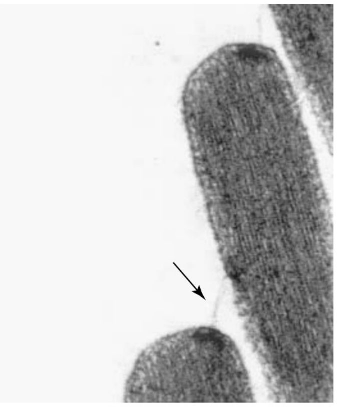


FIGURE 33.32 Electron micrograph of tip links. The tip link between two hair fibers is marked by an arrow. [Courtesy of Dr. A. Jacobs and Dr. A. J. Hudspeth.]

The presence of these tip links suggests a simple mechanical model for transduction by hair cells (Figure 33.33). The tip links are coupled to ion channels in the membranes of the stereocilia that are gated by mechanical stress. In the absence of a stimulus, approximately 15% of these channels are open. When the hair bundle is displaced toward its tallest part, the stereocilia slide across one another and the tension on the tip links increases, causing additional channels to open. The flow of ions through the newly opened channels depolarizes the membrane. Conversely, if the displacement is in the opposite direction, the tension on the tip links decreases, the open channels close, and the membrane hyperpolarizes. *Thus, the mechanical motion of the hair bundle is directly converted into current flow across the hair-cell membrane*.

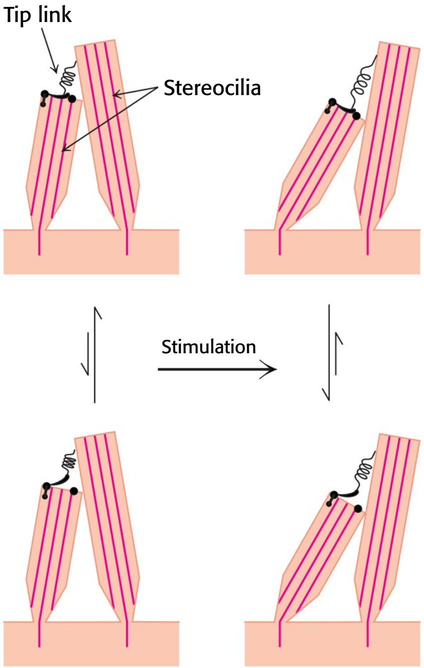


FIGURE 33.33 Model for hair-cell transduction. When the hair bundle is tipped toward the tallest part, the tip link pulls on an ion channel and opens it. Movement in the opposite direction relaxes the tension in the tip link, increasing the probability that any open channels will close. [Information from A. J. Hudspeth, *Nature* 341:397–404, 1989.]

Mechanosensory channels have been identified in *Drosophila* and vertebrates

The search for ion channels that respond to mechanical impulses has been pursued in a variety of organisms. *Drosophila* have sensory bristles used for detecting small air currents. These bristles respond to mechanical displacement in ways similar to those of hair cells; displacement of a bristle in one direction leads to substantial transmembrane current. Strains of mutant fruit flies that show uncoordinated motion and clumsiness have been examined for their electrophysiological responses to displacement of the sensory bristles. In one set of strains, transmembrane currents were dramatically reduced. The mutated gene in these strains was found to encode a protein of 1619 amino acids, called NompC for *no m*echanoreceptor *p*otential.

The carboxyl-terminal 469 amino acids of NompC resemble a class of ion-channel proteins called TRP (*t*ransient *r*eceptor *p*otential) channels. This region includes six putative transmembrane helices with a pore-like region between the fifth and the sixth helices. The amino-terminal 1150 amino acids consist almost exclusively of 29 *ankyrin repeats* (Figure 33.34).

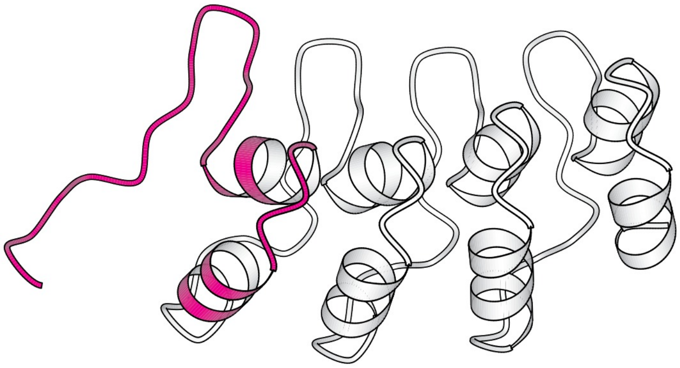


FIGURE 33.34 Ankyrin repeat structure. One ankyrin domain is shown in red in this series of four ankyrin repeats. *Notice* the hairpin loop followed by a helix-turn-helix motif in the red-colored ankyrin unit. Ankyrin domains interact with other proteins, primarily through their loops. [Drawn from 1AWC.pdb.]

Ankyrin repeats are structural motifs consisting of a hairpin loop followed by a helix-turn-helix. Importantly, in other proteins, regions with tandem arrays of these motifs mediate protein–protein interactions, suggesting that these arrays couple the motions of other proteins to the activity of the NompC channel.

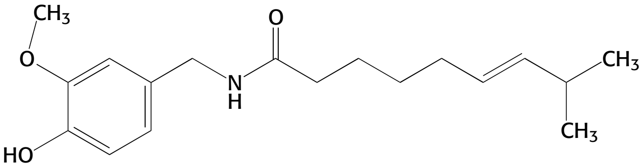
A candidate for at least one component of the mechanosensory channel taking part in hearing has been identified. The protein, TRPA1, is a member of the TRP channel family. The sequence of TRPA1 also includes 17 ankyrin repeats. TRPA1 is expressed in hair cells, particularly near their tips. Further studies are under way to test and extend the exciting hypothesis that this protein is a component of the sought-after channel.

33.5 Touch Includes the Sensing of Pressure, Temperature, and Other Factors

Like taste, touch is a combination of sensory systems that are expressed in a common organ—in this case, the skin. The detection of pressure and the detection of temperature are two key components. Amiloride-sensitive Na^+ channels, homologous to those of taste, appear to play a role. Other systems are responsible for detecting painful stimuli such as high temperature, acid, or certain specific chemicals. Although our understanding of this sensory system is not as advanced as that of the others, recent work has revealed a fascinating relation between pain and taste sensation, a relation well known to anyone who has eaten "spicy" food.

Studies of capsaicin reveal a receptor for sensing high temperatures and other painful stimuli

Our sense of touch is intimately connected with the sensation of pain. Specialized neurons, termed *nociceptors,* transmit signals from skin to pain-processing centers in the spinal cord and brain in response to the onset of tissue damage. What is the molecular basis for the sensation of pain? An intriguing clue came from the realization that *capsaicin,* the chemical responsible for the "hot" taste of spicy food, activates nociceptors.



Capsaicin

Early research suggested that capsaicin would act by opening ion channels that are expressed in nociceptors. Thus, a cell that expresses the *capsaicin receptor* should take up calcium on treatment with the molecule. This insight led to the isolation of the capsaicin receptor with the use of cDNA from cells expressing this receptor. Such cells had been detected by their fluorescence when loaded with the calcium-sensitive compound Fura-2 and then treated with capsaicin or related molecules. Cells expressing the capsaicin receptor, which is called VR1 (for *v*anilloid *r*eceptor 1), respond to capsaicin below a concentration of 1 μM. The deduced 838-residue sequence of VR1 revealed it to be a member of the TRP channel family (Figure 33.35). The amino-terminal region of VR1 includes three ankyrin repeats.

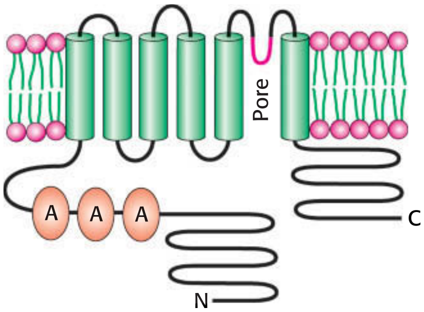


FIGURE 33.35 The membrane topology deduced for VR1, the capsaicin receptor. The proposed site of the membrane pore is indicated in red, and the three ankyrin (A) repeats are shown in orange. The active receptor comprises four of these subunits. [Information from M. J. Caterina et al., *Nature* 389:816–824, 1997.]

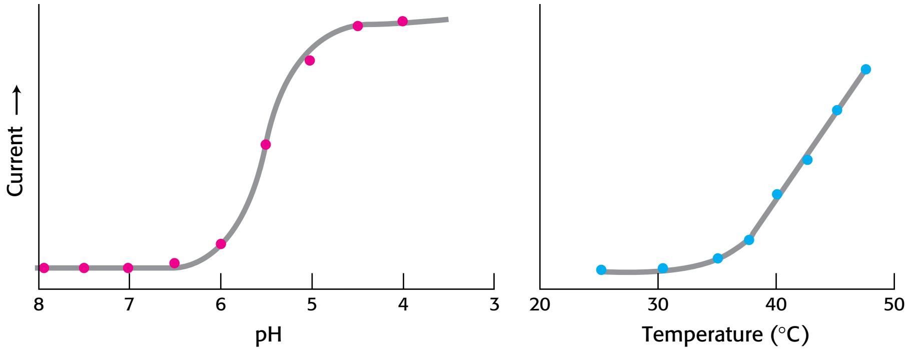


FIGURE 33.36 Response of the capsaicin receptor to pH and temperature. The ability of this receptor to respond to acid and to increased temperature helps detect potentially noxious situations. [Data from M. Tominaga et al., *Neuron* 21:531–543, 1998.]

Currents through VR1 are also induced by temperatures above 40°C and by exposure to dilute acid, with a midpoint for activation at pH 5.4 (Figure 33.36). Temperatures and acidity in these ranges are associated with infection and cell injury. The responses to capsaicin, temperature, and acidity are not independent. The response to heat is greater at lower pH, for example. Thus, *VR1 acts to integrate several noxious stimuli*. We feel these responses as pain and act to prevent the potentially destructive conditions that cause the unpleasant sensation. Studies of mice that do not express VR1 suggest that this is the case; such mice do not mind food containing high concentrations of capsaicin and are, indeed, less responsive than control mice to normally noxious heat. Plants such as chili peppers presumably gained the ability to synthesize capsaicin and other "hot" compounds to protect themselves from being consumed by mammals. Birds, which play the beneficial role of spreading pepper seeds into new territory, do not appear to respond to capsaicin.

Because of its ability to simulate VR1, capsaicin is used in pain management for arthritis, neuralgia, and other neuropathies. How can a compound that induces pain assist in its alleviation? Chronic exposure to capsaicin overstimulates pain-transmitting neurons, leading to their desensitization.

SUMMARY

Smell, taste, vision, hearing, and touch are based on signal-transduction pathways activated by signals from the environment. These sensory systems function similarly to the signal-transduction pathways for many hormones. These intercellular signaling pathways appear to have been appropriated and modified to process environmental information.

33.1 A Wide Variety of Organic Compounds Are Detected by Olfaction

The sense of smell, or olfaction, is remarkable in its specificity; it can, for example, discern stereoisomers of small organic compounds as distinct aromas. The 7TM receptors that detect these odorants operate in conjunction with $G_{(olf)}$, a G protein that activates a cAMP cascade resulting in the opening of an ion channel and the generation of a nerve impulse. An outstanding feature of the olfactory system is its ability to detect a vast array of odorants. Each olfactory neuron expresses only one type of receptor and connects to a particular region of the olfactory bulb. Odors are decoded by a combinatorial mechanism: each odorant activates a number of receptors, each to a different extent, and most receptors are activated by more than one odorant.

33.2 Taste Is a Combination of Senses That Function by Different Mechanisms

We can detect only five tastes: bitter, sweet, salt, sour, and umami. The transduction pathways that detect taste are, however, diverse. Bitter, sweet, and umami tastants are experienced through 7TM receptors acting through a special G protein called gustducin. Salty and sour tastants act directly through membrane channels. Salty tastants are detected by passage though Na^+ channels, whereas sour taste results from the effects of hydrogen ions on a number of types of channels. The end point is the same in all cases—membrane polarization that results in the transmission of a nerve impulse.

33.3 Photoreceptor Molecules in the Eye Detect Visible Light

Vision is perhaps the best understood of the senses. Two classes of photoreceptor cells exist: cones, which respond to bright lights and colors, and rods, which respond only to dim light. The photoreceptor in rods is rhodopsin, a 7TM receptor that is a complex of the protein opsin and the chromophore 11-*cis*-retinal. The absorption of light by 11-*cis*-retinal changes its structure into that of all-*trans*-retinal, setting in motion a signal-transduction pathway that leads to the breakdown of cGMP, to membrane hyperpolarization, and to a subsequent nerve impulse. Color vision is mediated by three distinct 7TM photoreceptors that employ 11-*cis*-retinal as a chromophore and absorb light in the blue, green, and red parts of the spectrum.

33.4 Hearing Depends on the Speedy Detection of Mechanical Stimuli

The immediate receptors for hearing are found in the hair cells of the cochleae, which contain bundles of stereocilia. When the stereocilia move in response to sound waves, cation channels will open or close, depending on the direction of movement. The mechanical motion of the cilia is converted into current flow and then into a nerve impulse.

33.5 Touch Includes the Sensing of Pressure, Temperature, and Other Factors

Touch, detected by the skin, senses pressure, temperature, and pain. Specialized nerve cells called nociceptors transmit signals that are interpreted in the brain as pain. A receptor responsible for the perception of pain has been isolated on the basis of its ability to bind capsaicin, the molecule responsible for the hot taste of spicy food. The capsaicin receptor, also called VR1, functions as a cation channel that initiates a nerve impulse.

KEY TERMS

main olfactory epithelium (p. 962)
$G_{(olf)}$ (p. 963)
odorant receptor (p. 963)
bitter (p. 966)
sweet (p. 966)
sour (p. 966)
salty (p. 966)
umami (p. 966)
gustducin (p. 967)
amiloride-sensitive Na^+ channel (p. 969)
rod (p. 970)
cone (p. 970)
rhodopsin (p. 970)
opsin (p. 970)
retinal (p. 970)
chromophore (p. 971)
transducin (p. 972)
cGMP phosphodiesterase (p. 972)
cGMP-gated ion channel (p. 972)
rhodopsin kinase (p. 972)
arrestin (p. 972)
guanylate cyclase (p. 973)
hair cell (p. 975)
stereocilia (p. 975)
tip link (p. 976)
nociceptor (p. 977)
capsaicin receptor (VR1) (p. 977)

PROBLEMS

1. *Olfaction in worms*. Unlike the olfactory neurons in the mammalian systems discussed herein, olfactory neurons in the nematode *C. elegans* express multiple olfactory receptors. In particular, one neuron (called AWA) expresses receptors for compounds to which the nematode is attracted, whereas a different neuron (called AWB) expresses receptors for compounds that the nematode avoids. Suppose that a transgenic nematode is generated such that one of the receptors for an attractant is expressed in AWB rather than AWA. What behavior would you expect in the presence of the corresponding attractant?

2. *Odorant matching*. A mixture of two of the compounds illustrated in Figure 33.6 is applied to a section of olfactory epithelium. Only receptors 3, 5, 9, 12, and 13 are activated, according to Figure 33.7. Identify the likely compounds in the mixture.

3. *Timing*. Compare the aspects of taste (bitter, sweet, salty, sour) in regard to their potential for rapid time resolution.

4. *Two ears*. Our ability to determine the direction from which a sound is coming is partly based on the difference in time at which our two ears detect the sound. Given the speed of sound (350 m s^{-1}) and the separation between our ears (0.15 m), what difference is expected in the times at which a sound arrives at our two ears? How does this difference compare with the time resolution of the human hearing system? Would a sensory system that utilized 7TM receptors and G proteins be capable of adequate time resolution?

5. *It's better to be bitter*. Some nontoxic plants taste very bitter to us. Suggest one or more explanations.

6. *Of mice and men*. In human beings, the umami taste is triggered only by glutamate and aspartate. In contrast, mice respond to many more amino acids. Design an experiment to test which of the subunits (T1R1 or T1R3) determines the specificity of this response. Assume that all desired mouse strains can be readily produced.

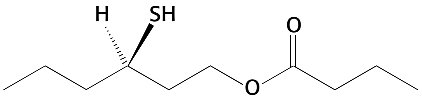
7. *It's not easy being green*. Would you expect light with wavelength 530 nm to be absorbed only by the "green" photoreceptor?

8. *Color discrimination*. Some women who have had male children who display color blindness have been demonstrated to distinguish between colors that are not readily distinguished by other people. Propose an explanation.

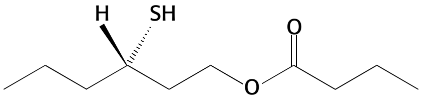
9. *Combinatorial power*. How many different odorants could a person distinguish if each odorant is bound by a single type of receptor and each receptor binds only a single odorant? How many if each odorant binds to two different odorant receptors? Three odorant receptors?

10. *Light action*. Describe the effect of light absorption on 11-*cis*-retinal bound within rhodopsin.

11. *Mirror, mirror*. Compounds A and B typically evoke different odors, with compound A described as smelling like tropical fruit and compound B smelling oniony.



A



B

Explain how two such similar compounds can smell so different.

12. *Channeling*. Provide an example of an ion channel in vision, taste, and hearing.

Chapter Integration Problem

13. *Energy and information*. The transmission of sensory information requires the input of free energy. For each sensory system (olfaction, gustation, vision, hearing, and touch), identify mechanisms for the input of free energy that allow the transmission of sensory information.

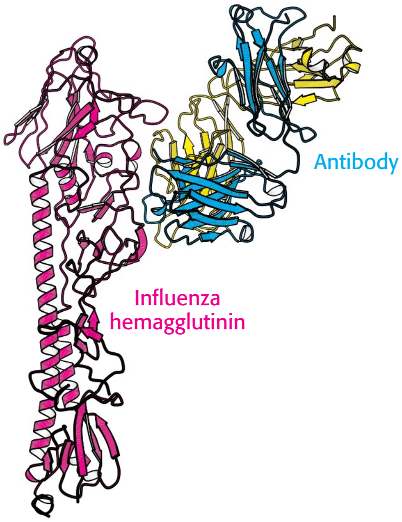
Mechanism Problem

14. *Schiff-base formation*. Propose a mechanism for the reaction between opsin and 11-*cis*-retinal.

CHAPTER

34

The Immune System



Just as medieval fortresses, such as the Dover Castle, used walls and fortifications to defend their territory, the immune system constantly battles against foreign invaders such as viruses, bacteria, and parasites to defend the organism. Antibody molecules provide a key element in the immune system's defensive arsenal. For example, specific antibodies can bind to molecules on the surfaces of viruses and prevent the viruses from infecting cells. Above right, an antibody binds to one subunit on hemagglutinin from the surface of influenza virus. [Skyscan Photolibrary/Alamy.]

We are constantly exposed to an incredible diversity of bacteria, viruses, and parasites, many of which would flourish in our cells or extracellular fluids were it not for our immune system. How does the immune system protect us? The human body has two lines of defense: an *innate immune system* that responds rapidly to features present in many pathogens, and an *adaptive immune system* that responds to specific features present only in a given pathogen. *Both the innate and the adaptive immune systems first identify features on disease-causing organisms and then work to eliminate or neutralize those organisms.* While a thorough description of the immune system is certainly beyond the scope of this book, this chapter will focus on how biochemical concepts such as protein structure, receptor-ligand interactions, and signal transduction are applied to the identification of pathogens.

The immune system must meet two tremendous challenges in the identification of pathogens: (1) to produce a system of receptors diverse enough to recognize a wide array of potential pathogens and (2) to distinguish invaders and their disease-causing products from the organism's own products (i.e., self- versus nonself-recognition). To meet these challenges, the innate immune system evolved the ability to recognize structural elements,

OUTLINE

34.1 Antibodies Possess Distinct Antigen-Binding and Effector Units

34.2 Antibodies Bind Specific Molecules Through Hypervariable Loops

34.3 Diversity Is Generated by Gene Rearrangements

34.4 Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors

34.5 The Immune System Contributes to the Prevention and the Development of Human Diseases

such as specific glycolipids or forms of nucleic acid, that are well conserved in pathogens but absent in the host organism. The repertoire of such elements is limited, however, and so some pathogens have strategies to escape detection. The adaptive immune system has the remarkable ability to produce more than 10^8 distinct proteins, called antibodies, that can recognize different foreign molecules, and more than 10^{12} receptors on immune cells, called T-cell receptors (TCRs), each of which presents a different surface with the potential to specifically bind a structure from a foreign organism. In producing this vast range of defensive molecules, however, the adaptive immune system has the potential to create antibodies and T-cells that recognize and attack cells or molecules normally present in our bodies—a situation that can result in autoimmune diseases.

This chapter will examine these challenges, focusing first on the structures of proteins that recognize foreign organisms and then on the mechanisms for protecting us from a specific pathogen once it has been recognized. The chapter will closely examine the modular construction of the proteins of the immune system—identifying structural motifs and considering how spectacular diversity can arise from modular construction.

Innate immunity is an evolutionarily ancient defense system

Innate immunity is an evolutionarily ancient defense system found, at least in some form, in all multicellular plants and animals. The innate immune system represents the first line of defense against foreign pathogens, relying on common features of invading organisms to identify and eliminate these threats. Components of the innate immune system include the epithelial lining that surrounds host cells and the specialized cells, called *phagocytes*, that can ingest and destroy pathogens without the aid of the adaptive immune system.

The innate immune system also includes a family of receptors that can recognize specific features present in most pathogens and yet not respond to materials normally present in the host. The best-understood of these receptors are the *Toll-like receptors* (TLRs). The name "toll-like" is derived from a receptor known as Toll encoded in the *Drosophila* genome; Toll was first identified in a screen for genes important for *Drosophila* development and was subsequently shown to play a key role in the innate immune system. The TLRs have a common structure (Figure 34.1). Each receptor consists of a large extracellular domain built primarily from repeated amino acid sequences termed leucine-rich repeats (LRRs). Each LRR typically contains 20–30 residues, including 6 that are usually leucine. The human TLRs have from 18 to 27 LRR repeats that are capped by a cysteine-rich domain and followed by a sequence forming a single transmembrane helix and an intracellular signaling domain. This signaling domain is not a protein kinase but acts as a docking site for other proteins. Most TLRs are expressed in the cell membrane for the detection of extracellular pathogens such as fungi and bacteria. Other TLRs are located in the membranes of internal compartments for the detection of intracellular pathogens such as viruses and some bacteria.

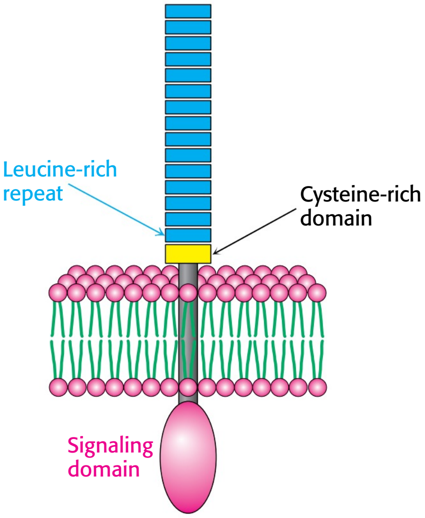


FIGURE 34.1 Toll-like receptor. Each receptor is made up of a set of 18 or more leucine-rich repeat sequences, followed by a cysteine-rich domain, a single transmembrane helix, and an intracellular domain that functions in signal transduction.

Each TLR targets a specific molecular characteristic, often called a *pathogen-associated molecular pattern* (PAMP), found primarily on invading organisms (Table 34.1). Typically, a PAMP is a critical component of the pathogen's function: mutations in these targets cannot easily block recognition by the TLR without compromising the activity of the pathogen. One particularly important PAMP is lipopolysaccharide (LPS), also referred to as *endotoxin*, a specific class of glycolipid found in the cell walls of Gram-negative bacteria such as *E. coli*. LPS is recognized by TLR4. The response

TABLE 34.1 Pathogen-associated molecular patterns (PAMPs) recognized by human TLRs

Receptor	PAMP	Source of PAMP
TLR1 (with TLR2)	Triacyl lipoprotein	Bacteria
TLR2	Lipoprotein	Bacteria, viruses, parasites
TLR3	Double-stranded RNA (dsRNA)	Viruses
TLR4	Lipopolysaccharide (LPS)	Bacteria, viruses
TLR5	Flagellin	Bacteria
TLR6 (with TLR2)	Diacyl lipoprotein	Bacteria, viruses
TLR7	Single-stranded RNA (ssRNA)	Viruses, bacteria
TLR8	Single-stranded RNA (ssRNA)	Viruses, bacteria
TLR9	CpG-DNA motifs	Viruses, bacteria, protozoa
TLR10	Unknown	Unknown

Data from: O. Takeuchi and S. Akira, *Cell* 140:805–820, 2010.

of the innate immune system to LPS can be easily demonstrated. Injection of less than 1 mg of LPS into a human being produces a fever and other signs of inflammation even though no living organisms are introduced.

How do TLRs recognize PAMPs? The extracellular domain from human TLR3 has a remarkable structure that is representative of many other TLRs (Figure 34.2). Each LRR contributes a single β strand to a large parallel β sheet that lines the inside of a concave, hooklike structure. In most cases, the PAMP is recognized by the surface formed between two receptors. The PAMP for TLR3, double-stranded RNA, fits in the groove between two monomers, colored blue and yellow in Figure 34.3. Ligand-induced dimerization of the receptor enables the initiation of a signaling cascade within the cell. We encountered similar examples of signal transduction across a membrane through receptor dimerization in Chapter 14.

Because the TLRs and other components of the innate immune system are always expressed, they provide the host organism with a rapid response to resist attack by pathogens. However, a number of pathogens have evolved the ability to escape detection by the innate immune system. For protection against such pathogens, the host relies on the adaptive immune system, which is able to target specific pathogens, even those that it has never encountered in the course of evolution.

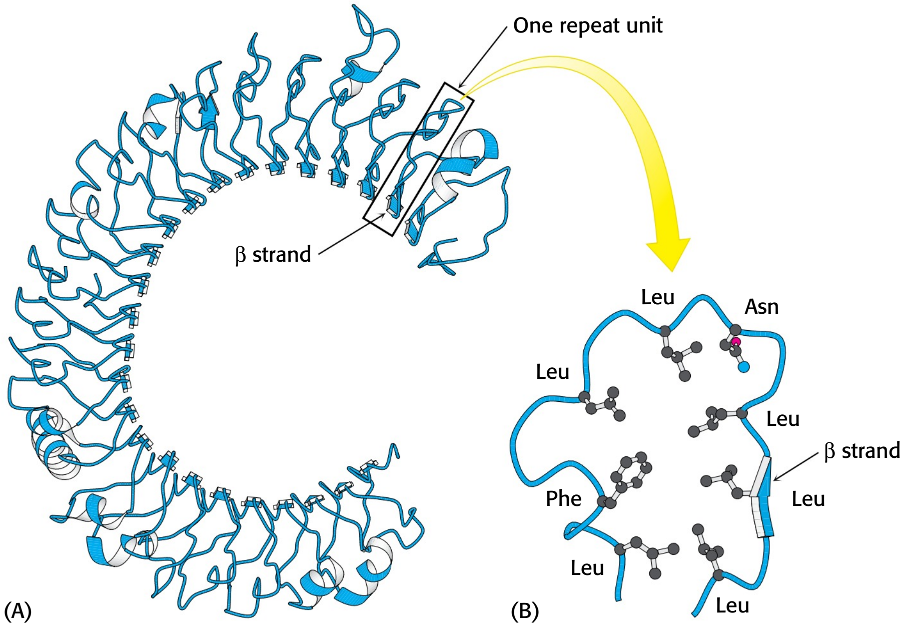


FIGURE 34.2 Extracellular domain of the Toll-like receptor. (A) The structure of the leucine-rich repeat (LRR) domain from human TLR3. *Notice* that the LRR units come together to form a central parallel β sheet that curls to form a concave structure. (B) The structure of a single LRR showing the positions of the residues that are generally approximately conserved. *Notice* that the leucine residues come together to form a hydrophobic core with the single β strand on one side. [Drawn from 1ZIW.pdb].

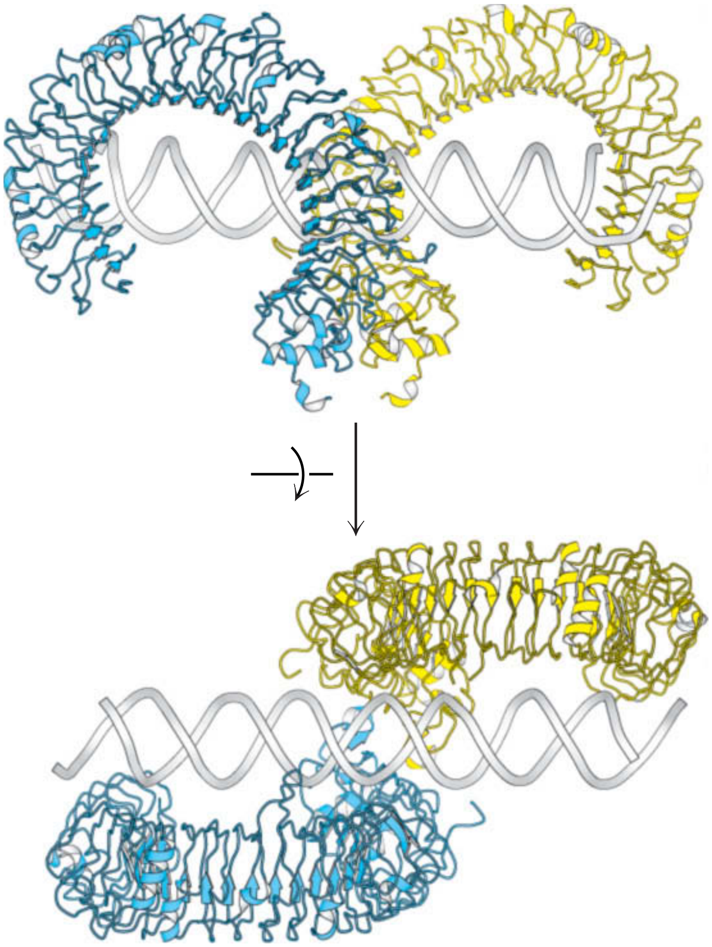


FIGURE 34.3 Recognition of a PAMP by a Toll-like receptor. The structure of TLR3 bound to its PAMP, a fragment of double-stranded RNA, as seen from the side (top) and from above (bottom). *Notice* that the PAMP induces receptor dimerization by binding the surfaces on the side of each of the extracellular domains. [Drawn from 3CIY.pdb].

The adaptive immune system responds by using the principles of evolution

The adaptive immune system comprises two parallel but interrelated systems: humoral and cellular immune responses. In the *humoral immune response,* soluble proteins called *antibodies (immunoglobulins)* function as recognition elements that bind to foreign molecules and serve as markers signaling foreign invasion. Antibodies are secreted by *plasma cells,* which are derived from *B lymphocytes (B cells)* (Figure 34.4). A foreign macromolecule that binds selectively to an antibody is called an *antigen.* In a physiological context, if the binding of the foreign molecule stimulates an immune response, that molecule is called an *immunogen*. The specific affinity of an antibody is not for the entire macromolecular antigen but for a particular site on the antigen called the *epitope* or *antigenic determinant.* Each B cell produces just one type of antibody that can recognize a single epitope.

In the *cellular immune response*, cells called *cytotoxic T lymphocytes* (also commonly called *killer T cells*) destroy cells that have been invaded by a pathogen. Because intracellular pathogens do not leave markings on the exteriors of infected cells, vertebrates have evolved a mechanism to mark the exterior of cells with a sample of the interior contents, both self and foreign. Some of the internal proteins are broken into peptides, which are then bound to a complex of integral membrane proteins encoded by the *major histocompatibility complex* (MHC). T cells continually scan the bound peptides to find and kill cells that display foreign motifs on their surfaces. Another class of T cells called *helper T lymphocytes (helper T cells)* contributes to both the humoral and the cellular immune responses by stimulating the differentiation and proliferation of appropriate B cells and cytotoxic T cells. The cellular immune response is mediated by specific receptors that are expressed on the surfaces of the T cells.

The remarkable ability of the immune system to adapt to an essentially limitless set of potential pathogens requires a powerful system for transforming the immune cells and molecules in response to the presence of pathogens. *This adaptive system operates through the principles of evolution, including reproduction with variation followed by selection of the most well suited members of a population.*

If the human genome contains, by the latest estimates, only 21,000 genes, how can the immune system generate more than 10^8 different antibody proteins and 10^{12} T-cell receptors? The answer is found in a novel mechanism for generating a highly diverse set of genes from a limited set of genetic building blocks. Linking different sets of DNA regions in a combinatorial manner produces many distinct protein-encoding genes that are not present in the genome. A rigorous selection process then leaves for proliferation only cells that synthesize proteins determined to be useful in the immune response. The subsequent reproduction of these cells without additional recombination enriches the cell population with members expressing particular protein species.

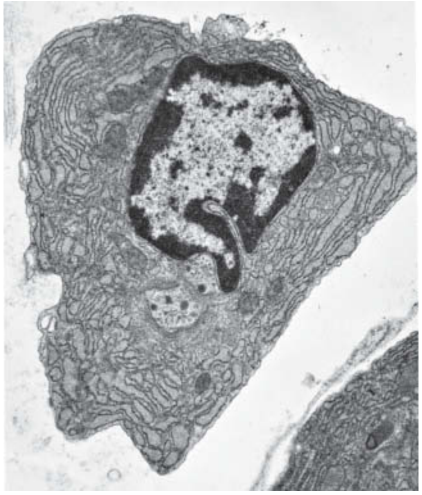


FIGURE 34.4 Immunoglobulin production. An electron micrograph of a plasma cell shows the highly developed rough endoplasmic reticulum necessary for antibody secretion. [Courtesy of Lynne Mercer.]

Critical to the development of the immune response is the selection process, which determines which immune cells will reproduce. The process comprises several stages. In the early stages of immune cell development, cells expressing molecules that bind tightly to self-molecules are destroyed or silenced, whereas cells expressing molecules that do not bind strongly to self-molecules and that have the potential for binding strongly to foreign molecules are preserved. The appearance of an immunogenic invader at a later time will stimulate the reproduction of cells expressing antibodies or T-cell receptors that bind specifically to elements of that pathogen—in evolutionary terms, such cells are positively selected. Thus, the immune response is based on the selection of cells expressing molecules that are specifically effective against a particular invader; the response evolves from a population with wide-ranging specificities to a more-focused collection of cells and molecules that are well suited to defend the host when confronted with that particular challenge.

Not only are antibodies and T-cell receptors a result of genetic diversity and recombination, but antibodies have highly diverse structures as well. Antibodies require many different structural solutions for binding many different antigens, each of which has a different form. T-cell receptors, in contrast, are not structurally diverse, because they have coevolved with the MHC. The docking mode of a T-cell receptor to the peptide bound to MHC is similar for all structures. As a consequence of this coevolution, every T-cell receptor has an inherent reactivity with every MHC. The coevolution ensures that all T-cell receptors can scan all peptide–MHC complexes on all tissues. The genetic diversity of the 10^{12} different T-cell receptors is concentrated in a highly diverse set of residues in the center of the MHC groove. This localized diversity allows the T-cell receptor to recognize the many different foreign peptides bound to the MHC. T-cell receptors must survey many different MHC–peptide complexes with rapid turnover. Therefore, the binding affinities between T-cell receptors and the MHC are weaker than those between antibody and antigen.

34.1 Antibodies Possess Distinct Antigen-Binding and Effector Units

Antibodies are central molecular players in the immune response. In Chapter 3, we discussed the application of antibodies as tools to study proteins; let us now consider the native structures and functions of these remarkable molecules. In 1959, Rodney Porter showed that *immunoglobulin G* (IgG), the major antibody in serum, can be cleaved into three 50-kDa fragments by the limited proteolytic action of papain. Two of these fragments bind antigen. They are called F_{ab} (*F* stands for *f*ragment, *ab* for *a*ntigen *b*inding). The other fragment, called F_c because it crystallizes readily, does not bind antigen, but it has other important biological activities, including the mediation of responses termed *effector functions*. These functions include the initiation of the *complement cascade,* a process that leads to the lysis of target cells. Although such effector functions are crucial to the functioning of the immune system, they will not be considered further here.

How do these fragments relate to the three-dimensional structure of whole IgG molecules? Immunoglobulin G consists of two kinds of polypeptide chains, a 25-kDa *light* (L) *chain* and a 50-kDa *heavy* (H) *chain*

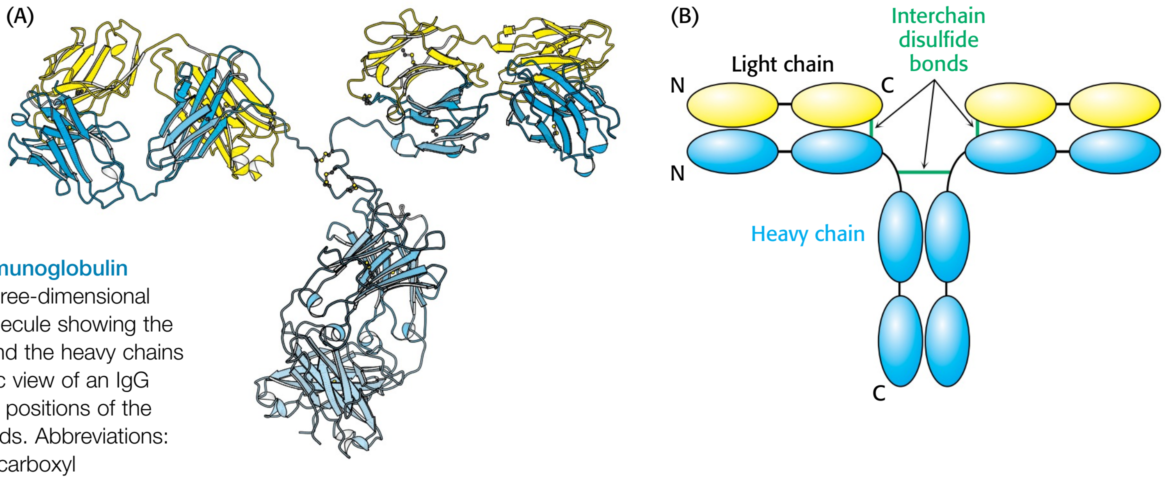


FIGURE 34.5 Immunoglobulin G structure. (A) The three-dimensional structure of an IgG molecule showing the light chains in yellow and the heavy chains in blue. (B) A schematic view of an IgG molecule indicating the positions of the interchain disulfide bonds. Abbreviations: N, amino terminus; C, carboxyl terminus. [Drawn from 1IGT.pdb.]

(Figure 34.5). The subunit composition is L_2H_2. Each L chain is linked to an H chain by a disulfide bond, and the H chains are linked to each other by at least one disulfide bond. Examination of the amino acid sequences and three-dimensional structures of IgG molecules reveals that each L chain comprises two homologous domains, termed *immunoglobulin domains,* to be described in detail in Section 34.2. Each H chain has four immunoglobulin domains. Overall, the molecule adopts a conformation that resembles the letter Y. The stem of this Y, corresponding to the F_c fragment obtained by cleavage with papain, consists of the two carboxyl-terminal immunoglobulin domains of each H chain. The two arms of the Y, corresponding to the two F_{ab} fragments, are formed by the two amino-terminal domains of each H chain and the two domains of each L chain. The linkers between the stem and the two arms consist of extended polypeptide regions within the H chains and are quite flexible.

Papain cleaves the H chains on the carboxyl-terminal side of the disulfide bond that links each L and H chain (Figure 34.6). Thus, each F_{ab} consists of an entire L chain and the amino-terminal half of an H chain, whereas F_c consists of the carboxyl-terminal halves of both H chains. Each F_{ab} contains a single antigen-binding site. Because an intact IgG molecule contains two F_{ab} components and therefore has two binding sites, it can cross-link multiple antigens (Figure 34.7). Furthermore, the F_c and the two F_{ab} units of the intact IgG are joined by flexible polypeptide regions that allow

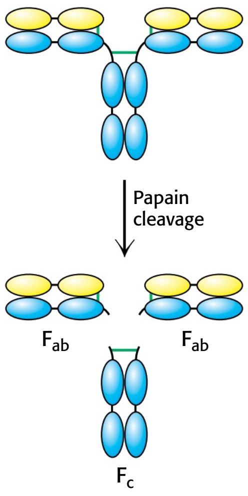


FIGURE 34.6 Immunoglobulin G cleavage. Treatment of intact IgG molecules with the protease papain results in the formation of three large fragments: two F_{ab} fragments that retain antigen-binding capability and one F_c fragment that does not.

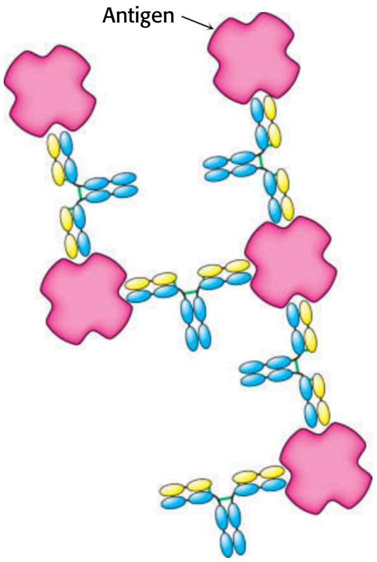


FIGURE 34.7 Antigen cross-linking. Because IgG molecules include two antigen-binding sites, antibodies can cross-link multivalent antigens such as viral surfaces.

TABLE 34.2 Properties of immunoglobulin classes

Class	Serum concentration (mg ml^{-1})	Mass (kDa)	Light chains	Heavy chains	Chain structure
IgG	12	150	κ or λ	γ	$\kappa_2\gamma_2$ or $\lambda_2\gamma_2$
IgA	3	180–500	κ or λ	α	$(\kappa_2\alpha_2)_n$ or $(\lambda_2\alpha_2)_n$
IgM	1	950	κ or λ	μ	$(\kappa_2\mu_2)_5$ or $(\lambda_2\mu_2)_5$
IgD	0.1	175	κ or λ	δ	$\kappa_2\delta_2$ or $\lambda_2\delta_2$
IgE	0.001	200	κ or λ	ε	$\kappa_2\varepsilon_2$ or $\lambda_2\varepsilon_2$

Note: n = 1, 2, or 3. IgM and oligomers of IgA also contain J chains that connect immunoglobulin molecules. IgA in secretions has an additional component.

facile variation in the angle between the F_{ab} units through a wide range (Figure 34.8). This kind of mobility, called *segmental flexibility,* can enhance the formation of an antibody–antigen complex by enabling both recognition sites on an antibody to bind an antigen that possesses multiple binding sites, such as a viral coat composed of repeating identical monomers. The sites at the tips of the F_{ab} units simply move to match the distance between specific determinants on the antigen.

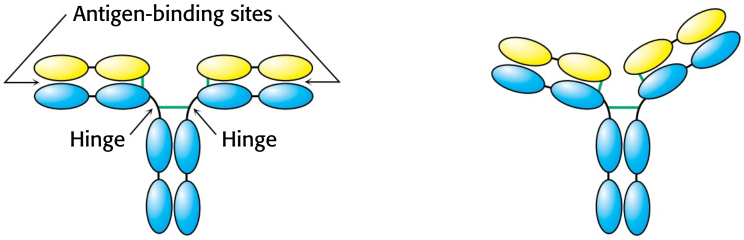
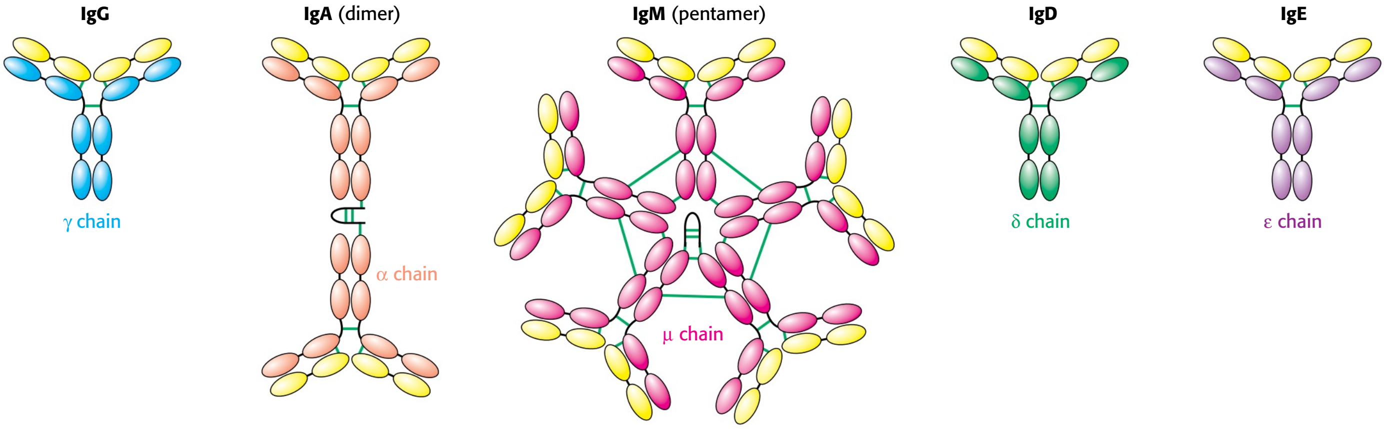


FIGURE 34.8 Segmental flexibility. The linkages between the F_{ab} and the F_c regions of an IgG molecule are flexible, allowing the two antigen-binding sites to adopt a range of orientations with respect to each other. This flexibility allows effective interactions with a multivalent antigen without requiring that the epitopes on the target be a precise distance apart.

Immunoglobulin G is the antibody present in highest concentration in the serum, but other classes of immunoglobulin also are present (Table 34.2). Each class includes an L chain (either κ or λ) and a distinct H chain (Figure 34.9). The heavy chains in IgG are called γ chains, whereas those in immunoglobulins A, M, D, and E are called α, μ, δ, and ε, respectively. *Immunoglobulin M* (IgM) is the first class of antibody to appear in the serum after exposure to an antigen. The presence of 10 antigen recognition sites enables IgM to bind especially tightly to antigens containing multiple identical epitopes. The strength of an interaction comprising multiple independent binding interactions between partners is termed *avidity* rather than affinity, which denotes the binding strength of a single binding site.

Immunoglobulin A (IgA) is the major class of antibody in external secretions, such as saliva, tears, bronchial mucus, and intestinal mucus. Thus, IgA serves as a first line of defense against bacterial and viral antigens. The role of *immunoglobulin D* (IgD) has long been a mystery, but recent studies suggest IgD plays a role in the activation of basophils, white blood cells that have antiparasitic functions. *Immunoglobulin E* (IgE) is important in conferring

FIGURE 34.9 Classes of immunoglobulin. Each of five classes of immunoglobulin has the same light chain (shown in yellow) combined with a different heavy chain (γ, α, μ, δ, or ε). Disulfide bonds are indicated by green lines. The IgA dimer and the IgM pentamer have a small polypeptide chain in addition to the light and heavy chains.



protection against parasites, but IgE also participates in allergic reactions. IgE–antigen complexes form cross-links with receptors on the surfaces of mast cells to trigger a cascade that leads to the release of granules containing pharmacologically active molecules. Histamine, one of the agents released, induces smooth-muscle contraction and stimulates the secretion of mucus.

34.2 Antibodies Bind Specific Molecules Through Hypervariable Loops

A comparison of the amino acid sequences of different IgG antibodies from human beings or mice shows that the carboxyl-terminal half of the L chains and the carboxyl-terminal three-quarters of the H chains are very similar in all of the antibodies. Importantly, the amino-terminal domain of each chain is more variable, including three stretches of approximately 7 to 12 amino acids within each chain that are hypervariable, as shown for the H chain in Figure 34.10. The amino-terminal immunoglobulin domain of each chain is thus referred to as the *variable region,* whereas the remaining immunoglobulin domains are much more similar in all antibodies and are referred to as *constant regions* (Figure 34.11).

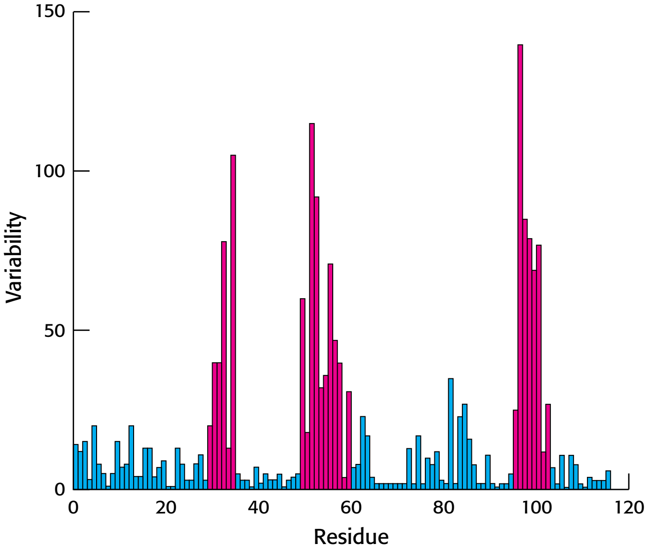


FIGURE 34.10 Immunoglobulin sequence diversity. A plot of sequence variability as a function of position along the sequence of the amino-terminal immunoglobulin domain of the H chain of human IgG molecules. Three regions (in red) show remarkably high levels of variability. These hypervariable regions correspond to three loops in the immunoglobulin domain structure. [Data from R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology*, 4th ed. (W. H. Freeman and Company, 2000), p. 91.]

The immunoglobulin fold consists of a beta-sandwich framework with hypervariable loops

An IgG molecule consists of a total of 12 immunoglobulin domains. These domains have many sequence features in common and adopt a common structure, the *immunoglobulin fold* (Figure 34.12). Remarkably, this same structural domain is found in many other proteins that play key roles in both immune and nonimmune functions.

The immunoglobulin fold consists of a pair of β sheets, each built of antiparallel β strands, that surround a central hydrophobic core. A single disulfide bond bridges the two sheets. Two aspects of this structure are particularly important for its function. First, three loops present at one end of the structure form a potential binding surface. These loops contain the hypervariable sequences present in antibodies and in T-cell receptors. Variation of the amino acid sequences of these loops provides the major mechanism for the generation of the vastly diverse set of antibodies and T-cell receptors expressed by the immune system. These loops are referred to as *hypervariable loops* or *complementarity-determining regions* (CDRs). Second, the amino terminus and the carboxyl terminus are at opposite ends of the structure, which allows structural domains to be strung together to form chains, as in the L and H chains of antibodies. Such chains are present in several other key molecules in the immune system.

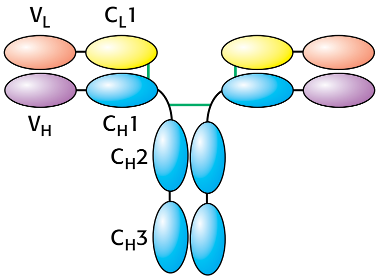


FIGURE 34.11 Variable and constant regions. Each L and H chain includes one immunoglobulin domain at its amino terminus that is quite variable from one antibody to another. These domains are referred to as V_L and V_H. The remaining domains are more constant from one antibody to another and are referred to as constant domains (C_L1, C_H1, C_H2, and C_H3).

The immunoglobulin fold is one of the most prevalent domains encoded by the human genome: more than 750 genes encode proteins with at least one immunoglobulin fold recognizable at the level of amino acid sequence. Such domains are also common in other multicellular animals such as flies and nematodes. However, from inspection of amino acid sequence alone, immunoglobulin-fold domains do not appear to be present in yeast or plants, although these organisms possess other structurally similar domains, including the key photosynthetic electron-transport protein plastocyanin in plants (Section 19.3). Thus, the immunoglobulin-fold family appears to have expanded greatly along evolutionary branches leading to animals—particularly, vertebrates.

X-ray analyses have revealed how antibodies bind antigens

For each class of antibody, the variable domains at the amino-terminal ends of the L and H chains (designated V_L and V_H) come together to form a binding surface. The positions of the CDRs are striking. These hypervariable sequences, present in three loops of each domain, come together so that all six loops form a single surface at the end of each arm (Figure 34.13). Because virtually any V_L can pair with any V_H, *a very large number of different binding sites can be constructed by their combinatorial association.*

The results of x-ray crystallographic studies of several hundred large and small antigens bound to F_{ab} molecules have been sources of much insight into the structural basis of antibody specificity. The binding of antigens to antibodies is governed by the same principles that govern the binding of substrates to enzymes. The interaction between complementary shapes results in numerous contacts between amino acids at the binding surfaces of both molecules. Many hydrogen bonds, electrostatic interactions, and van der Waals interactions, reinforced by hydrophobic interactions, combine to give specific and strong binding.

A few aspects of antibody binding merit specific attention, inasmuch as they relate directly to the structure of immunoglobulins. The binding site on the antibody incorporates some or all of the CDRs in the variable domains of the antibody. Small molecules are likely to make contact with fewer CDRs, with perhaps 15 residues of the antibody participating in the binding interaction. Macromolecules often make more extensive contact, sometimes interacting with all six CDRs and 20 or more residues of the antibody. Small molecules often bind in a cleft of the antigen-binding region. Macromolecules, such as globular proteins, tend to interact across larger, fairly flat apposed surfaces bearing complementary protrusions and depressions.

A well-studied case of small-molecule binding is seen in an example of phosphorylcholine bound to F_{ab}. Crystallographic analysis revealed phosphorylcholine bound to a cavity lined by residues from five CDRs—two from the L chain and three from the H chain (Figure 34.14). The positively charged trimethylammonium group of phosphorylcholine is buried inside the wedge-shaped cavity, where it interacts electrostatically with two negatively charged residues, a glutamate and an aspartate. The negatively charged phosphoryl group of phosphorylcholine binds to the positively charged guanidinium group of an arginine residue at the mouth of the crevice and is hydrogen bonded to the side chain of a nearby tyrosine residue.

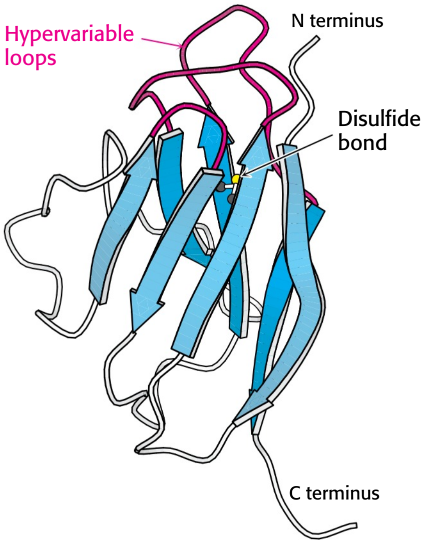


FIGURE 34.12 Immunoglobulin fold. An immunoglobulin domain consists of a pair of β sheets linked by a disulfide bond and hydrophobic interactions. *Notice* that three hypervariable loops lie at one end of the structure. [Drawn from 1DQJ.pdb.]

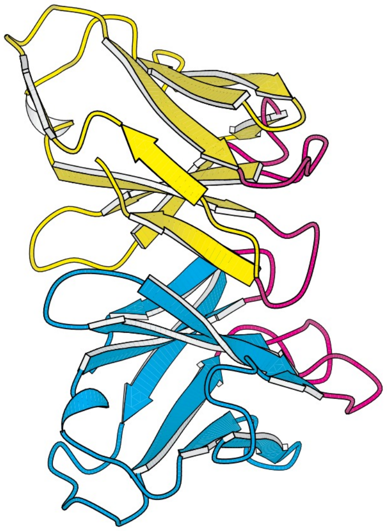


FIGURE 34.13 Variable domains. A side view of the variable domains of the L chain (yellow) and the H chain (blue); the complementarity-determining regions (CDRs) are shown in red. *Notice* that the six CDRs come together to form a binding surface. The specificity of the surface is determined by the sequences and structures of the CDRs. [Drawn from 1DQJ.pdb.]

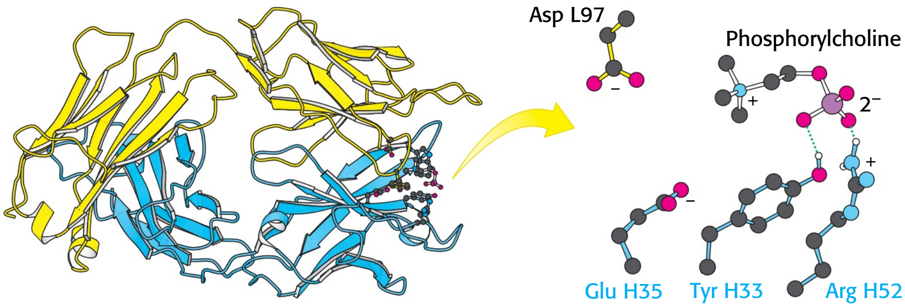


FIGURE 34.14 Binding of a small antigen. The structure of a complex between an F_{ab} fragment of an antibody (blue and yellow) and its target—in this case, phosphorylcholine. *Notice* that residues from the antibody interact with phosphorylcholine through hydrogen bonding and electrostatic interactions. [Drawn from 2MCP.pdb.]

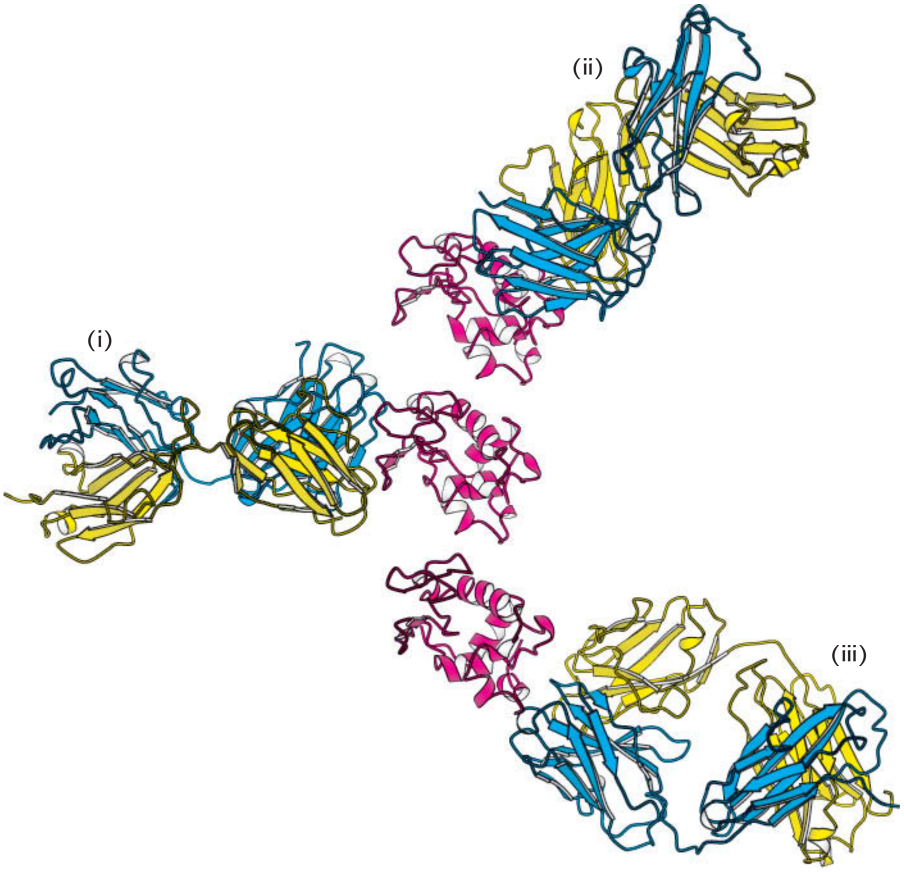


FIGURE 34.15 Antibodies against lysozyme. The structures of three complexes (i, ii, iii) between F_{ab} fragments (blue and yellow) and hen egg-white lysozyme (red) shown with lysozyme in the same orientation in each case. *Notice* that the three antibodies recognize completely different epitopes on the lysozyme molecule. [Drawn from 3HFL, 1DQJ, and 1FDL.pdb.]

Numerous van der Waals interactions, such as those made by a tryptophan side chain, also stabilize this complex.

Residues from five CDRs participate in the binding of phosphorylcholine to human F_{ab}. This binding does not significantly change the structure of the antibody, yet induced fit plays a role in the formation of many antibody–antigen complexes. A malleable binding site can accommodate many more kinds of ligands than can a rigid one. Thus, induced fit increases the repertoire of antibody specificities.

Large antigens bind antibodies with numerous interactions

How do large antigens interact with antibodies? A large collection of antibodies against hen egg-white lysozyme has been structurally characterized in great detail (Figure 34.15). Each different antibody binds to a distinct surface of lysozyme. Let us examine the interactions in one of these complexes (complex ii in Figure 34.15) in detail. This antibody binds two polypeptide segments of lysozyme that are widely separated in the primary structure (Figure 34.16).

All six CDRs of the antibody make contact with this epitope. The region of contact is quite extensive (about 30 × 20 Å). The apposed surfaces are rather flat. The only exception is the side chain of glutamine 121 of lysozyme, which penetrates deeply into the antibody's binding site, where it forms a hydrogen bond with a main-chain carbonyl oxygen atom and is surrounded by three aromatic side chains. The formation of 12 hydrogen bonds and numerous van der Waals interactions contributes to the high affinity ($K_d = 20$ nM) of this antibody–antigen interaction. Examination of the F_{ab} molecule without bound protein reveals that the structures of the V_L and V_H domains change little on binding, although they slide 1 Å apart to allow more intimate contact with lysozyme.

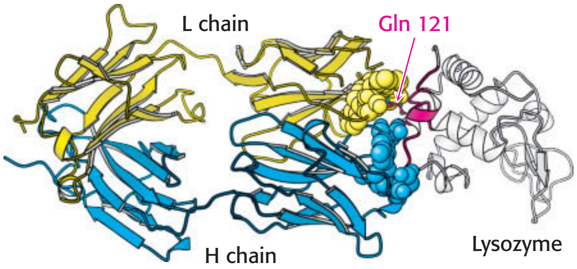


FIGURE 34.16 Antibody–protein interactions. *Notice* that a single residue of lysozyme, glutamine 121, penetrates more deeply into the antibody binding site. [Drawn from 1FDL.pdb.]

34.3 Diversity Is Generated by Gene Rearrangements

A mammal such as a mouse or a human being can synthesize large amounts of specific antibody against virtually any foreign determinant within a matter of days after exposure. We have seen that antibody specificity is determined by the amino acid sequences of the variable regions of both light and heavy chains, which brings us to the key question: How are different variable-region sequences generated?

The discovery of distinct variable and constant regions in the L and H chains raised the possibility that the genes that encode immunoglobulins have an unusual architecture that facilitates the generation of a diverse set of polypeptide products. In 1965, William Dreyer and Claude Bennett proposed that multiple *V (variable) genes* are separate from a single *C (constant) gene* in embryonic (germ-line) DNA. According to their model, one of these V genes becomes joined to the C gene in the course of differentiation of the antibody-producing cell. A critical test of this novel hypothesis had to await the isolation of pure immunoglobulin mRNA and the development of techniques for analyzing mammalian genomes. Twenty years later, Susumu Tonegawa found that V and C genes are indeed far apart in embryonic DNA but are closely associated in the DNA of antibody-producing cells. Thus, immunoglobulin genes are rearranged in the differentiation of lymphocytes.

J (joining) genes and D (diversity) genes increase antibody diversity

Sequencing studies carried out by Susumu Tonegawa, Philip Leder, and Leroy Hood revealed that V genes in embryonic cells do not encode the entire variable region of L and H chains. Consider, for example, the region that encodes the κ light-chain family. A tandem array of 40 segments, each of which encodes approximately the first 97 residues of the variable domain of the L chain, is present on human chromosome 2 (Figure 34.17).

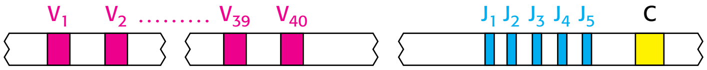


FIGURE 34.17 The κ light-chain locus. This part of human chromosome 2 includes an array of 40 segments that encode the variable (V) region (approximately residues 1–97) of the light chain, an array of 5 segments that encode the joining (J) region (residues 98–110), and a single region that encodes the constant (C) region.

However, the variable region of the L chain contains 110 residues. Where is the DNA that encodes the last 13 residues of the variable domain? For L chains in undifferentiated cells, this stretch of DNA is located in an unexpected place: near the C gene. It is called the *J gene* because it *j*oins the V and C gene segments in a differentiated cell. In fact, a tandem array of five J gene segments is located near the C gene in embryonic cells. In the differentiation of an antibody-producing cell, a V gene segment is joined to a J gene segment to form a complete gene for the variable region (Figure 34.18). RNA splicing (Section 29.3) generates an mRNA molecule for the complete

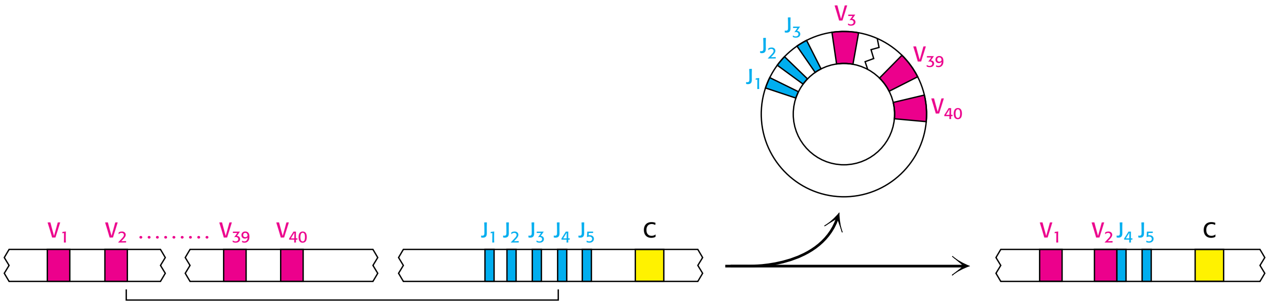


FIGURE 34.18 VJ recombination. A single V gene (in this case, V_2) is linked to a J gene (here, J_4) to form an intact VJ region. The intervening DNA is released in a circular form. Because the V and J regions are selected at random and the joint between them is not always in exactly the same place, many VJ combinations can be generated by this process.

L chain by linking the coding regions for the rearranged VJ unit with that for the C unit (Figure 34.19).

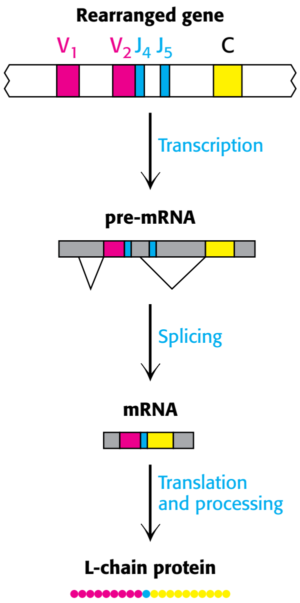


FIGURE 34.19 Light-chain expression. The light-chain protein is expressed by the transcription of the rearranged gene to produce a pre-RNA molecule with the VJ and C regions separated. RNA splicing removes the intervening sequences to produce an mRNA molecule with the VJ and C regions linked. Translation of the mRNA and processing of the initial protein product produce the light chain.

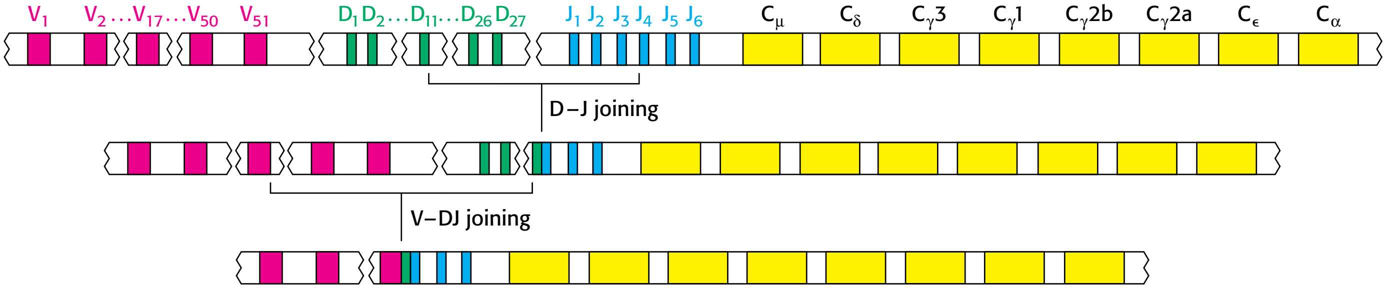
J gene segments are important contributors to antibody diversity because they encode part of the last hypervariable segment (CDR3). In forming a continuous variable-region gene, any of the 40 V gene segments can become linked to any of 5 J gene segments. Thus, somatic recombination of these gene segments amplifies the diversity already present in the germ line. The linkage between V and J is not precisely controlled. Recombination between these genes can take place at one of several bases near the codon for residue 95, generating additional diversity. A similar array of V and J gene segments encoding the λ light chain is present on human chromosome 22. This region includes 30 Vλ gene segments and four Jλ gene segments. In addition, this region includes four distinct C genes, in contrast with the single C gene in the κ locus.

In humans, the genes encoding the heavy chain are present on chromosome 14. Remarkably, the variable domain of heavy chains is assembled from *three* rather than two segments. In addition to V_H gene segments that encode residues 1 through 94 and J_H gene segments that encode residues 98 through 113, this chromosomal region includes a distinct set of segments that encode residues 95 through 97 (Figure 34.20). These gene segments are called D (for *d*iversity). Some 27 D gene segments lie between 51 V_H and 6 J_H gene segments. The recombination process first joins a D gene segment to a J_H gene segment; a V_H gene segment is then joined to DJ_H. A greater variety of antigen-binding patches and clefts can be formed by the H chain than by the L chain because the H chain is encoded by three rather than two gene segments. Moreover, CDR3 of the H chain is diversified by the action of terminal deoxyribonucleotidyl transferase, a special DNA polymerase that requires no template. This enzyme inserts extra nucleotides between V_H and D. The *V(D)J recombination* of both the L and the H chains is executed by specific enzymes present in immune cells. These proteins, called *RAG-1* and *RAG-2*, recognize specific DNA sequences called *recombination signal sequences* (RSSs) adjacent to the V, D, and J gene segments and facilitate the cleavage and religation of the DNA segments.

More than 10^8 antibodies can be formed by combinatorial association and somatic mutation

Let us recapitulate the sources of antibody diversity. The germ line contains a rather large repertoire of variable-region genes. For κ light chains, there are about 40 V-segment genes and 5 J-segment genes. Hence, a total of $40 \times 5 = 200$ kinds of complete V_κ genes can be formed by the combinations of V and J. A similar analysis suggests that at least 120 different λ light chains can be generated. A larger number of heavy-chain genes can be formed because of the role of the D segments. For 51 V, 27 D, and 6 J gene segments, the number of complete V_H genes that can be formed is 8262. The association of 320 kinds of L chains with 8262 kinds of H chains would yield 2.6×10^6

FIGURE 34.20 V(D)J recombination. The heavy-chain locus includes an array of 51 V segments, 27 D segments, and 6 J segments. Gene rearrangement begins with D–J joining, followed by further rearrangement to link the V segment to the DJ segment.



different antibodies. Variability in the exact points of segment joining and other mechanisms increases this value by at least two orders of magnitude.

Even more diversity is introduced into antibody chains by *somatic mutation*—that is, by the introduction of mutations into the recombined genes. In fact, a 1000-fold increase in binding affinity is seen in the course of a typical humoral immune response, arising from somatic mutation, a process called *affinity maturation*. The generation of an expanded repertoire leads to the selection of antibodies that more precisely fit the antigen. Thus, nature draws on each of three sources of diversity—a germ-line repertoire, somatic recombination, and somatic mutation—to form the rich variety of antibodies that protect an organism from foreign incursions.

The oligomerization of antibodies expressed on the surfaces of immature B cells triggers antibody secretion

The processes heretofore described generate a highly diverse set of antibody molecules—a key first step in the generation of an immune response. The next stage is the selection of a particular set of antibodies directed against a specific invader. How is this selection accomplished? Each immature B cell, produced in the bone marrow, expresses a specific monomeric form of IgM attached to its surface (Figure 34.21). Each cell expresses approximately 10^5 IgM molecules, but *all of these molecules are identical in amino acid sequence and, hence, in antigen-binding specificity*. Thus, the selection of a particular immature B cell for growth will lead to the amplification of an antibody with a unique specificity. The selection process begins with the binding of an antigen to the membrane-bound antibody.

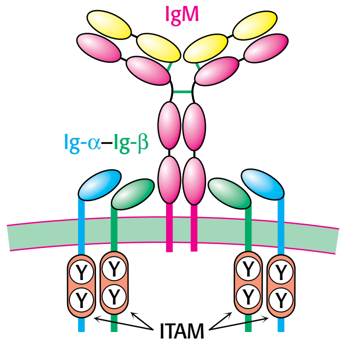


FIGURE 34.21 B-cell receptor. This complex consists of a membrane-bound IgM molecule noncovalently bound to two Ig-α–Ig-β heterodimers. The intracellular domains of each of the Ig-α and Ig-β chains include an immunoreceptor tyrosine-based activation motif (ITAM).

Associated with each membrane-linked IgM molecule are two molecules of a heterodimeric membrane protein called Ig-α–Ig-β (Figure 34.21). Examination of the amino acid sequences of Ig-α and Ig-β is highly instructive. The amino terminus of each protein lies outside the cell and corresponds to a single immunoglobulin fold, and the carboxyl terminus, which lies inside the cell, includes a sequence of 18 amino acids called an *immunoreceptor tyrosine-based activation motif* (ITAM; Figure 34.21). As its name suggests, each ITAM includes key tyrosine residues, which are subject to phosphorylation by particular protein kinases present in immune-system cells.

A fundamental observation linking the binding of antigen to membrane-bound antibody with the subsequent initiation of an immune response is that *oligomerization, or clustering, of the antibody molecules is required* (Figure 34.22). The requirement for oligomerization is reminiscent of the dimerization of receptors triggered by epidermal growth factor and insulin encountered in Section 14.3; indeed, the associated signaling mechanisms appear to be quite similar. The oligomerization of the membrane-bound antibodies results in the phosphorylation of the tyrosine residues within the ITAMs by protein tyrosine kinases including Lyn, a homolog of Src (Section 14.5). The phosphorylated ITAMs serve as docking sites for a protein kinase termed *s*pleen *t*yrosine *k*inase (Syk), which has two SH2 domains that interact with the pair of phosphorylated tyrosine residues in each ITAM. Docked Syk adopts an active confirmation, which then phosphorylates other signal-transduction proteins, including an inhibitory subunit of the transcription factor NF-κB and

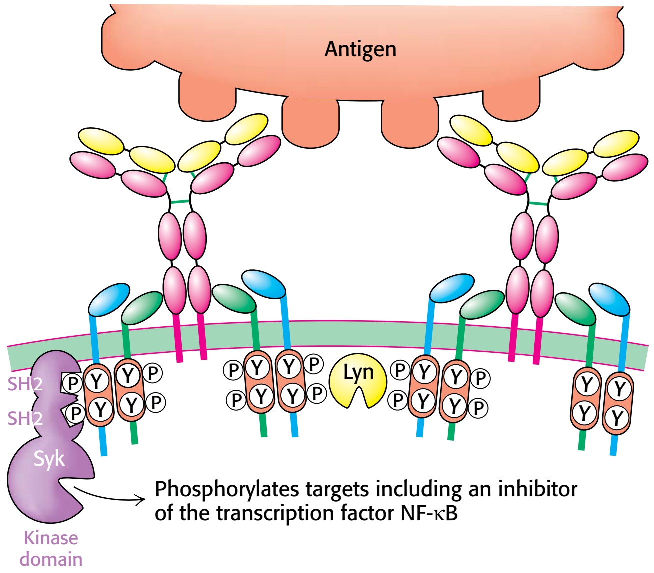
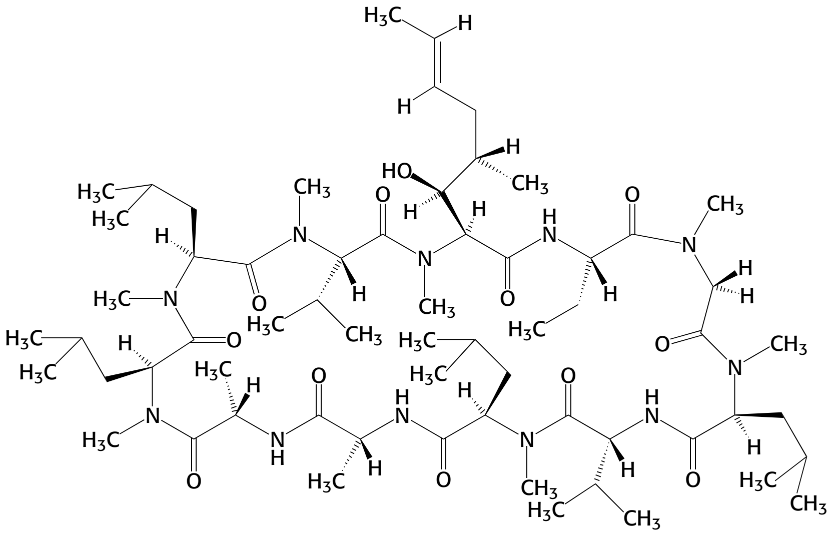

Antigen
SH2
SH2
P
Y
Lyn
Syk
Kinase domain
Phosphorylates targets including an inhibitor of the transcription factor NF-κB

FIGURE 34.22 B-cell activation. The binding of a multivalent antigen such as a bacterial or viral surface links membrane-bound IgM molecules. This oligomerization triggers the phosphorylation of tyrosine residues in the ITAM sequences by protein tyrosine kinases such as Lyn. After phosphorylation, the ITAMs serve as docking sites for Syk, a protein kinase that phosphorylates a number of targets, including transcription factors.

an isoform of phospholipase C. The signaling processes continue downstream to activate gene expression, leading to the stimulation of cell growth and initiating further B-cell differentiation.

Drugs that modulate the immune system have served as sources of insight into immune-system signaling pathways. For example, *cyclosporin,* a powerful suppressor of the immune system, inhibits *calcineurin,* a phosphatase which normally activates a transcription factor called NF-AT by dephosphorylating it.



Cyclosporin A

The potent inhibition of the immune system, or *immunosuppression*, induced by cyclosporin reveals how crucial the activity of NF-AT is to the development of an immune response. Without such drugs, organ transplantation would be extremely difficult because transplanted tissue expresses a wide range of foreign antigens, causing the host immune system to reject the new tissue.

The role of oligomerization in the B-cell signaling pathway is illuminated when we consider the nature of many antigens presented by pathogens. The surfaces of many viruses, bacteria, and parasites are characterized by arrays of identical membrane proteins or membrane-linked carbohydrates. Thus, most pathogens present multiple binding surfaces that will naturally cause membrane-associated antibodies to oligomerize as they bind adjacent epitopes. In addition, the mechanism accounts for the observation that most small molecules do not induce an immune response; however, coupling multiple copies of a small molecule to a large oligomeric protein such as keyhole limpet hemocyanin (KLH), which has a molecular mass of 1 million daltons or more, promotes antibody oligomerization and, hence, the production of antibodies against the small-molecule epitope. The large protein is called the *carrier* of the attached chemical group, which is called a *haptenic determinant*. The small foreign molecule by itself is called a *hapten*. Antibodies elicited by attached haptens will bind unattached haptens as well.

Different classes of antibodies are formed by the hopping of V_H genes

The development of an effective antibody-based immune response depends on the secretion into the blood of antibodies that have appropriate effector functions. At the beginning of this response, an alternative mRNA-splicing pathway is activated so that the production of membrane-linked IgM is

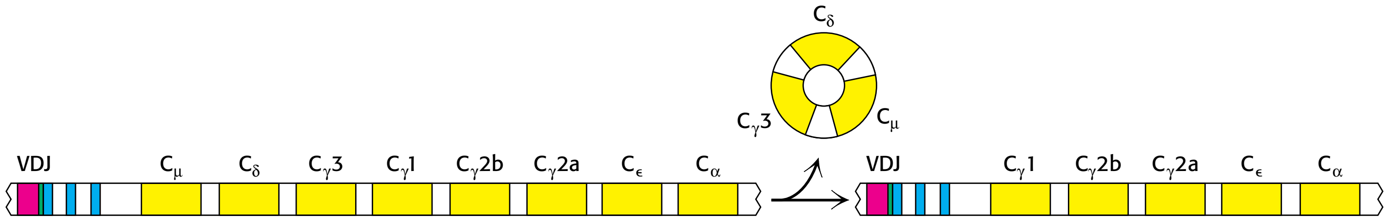


FIGURE 34.23 Class switching. Further rearrangement of the heavy-chain locus results in the generation of genes for antibody classes other than IgM. In the case shown, rearrangement places the VDJ region next to the Cγ1 region, resulting in the production of IgG1. Note that no further rearrangement of the VDJ region takes place, and so the specificity of the antibody is not affected.

supplanted by the synthesis of secreted IgM. As noted in Section 34.1, secreted IgM is pentameric and has a high avidity for antigens containing multiple identical epitopes. Later, the antibody-producing cell makes either IgG, IgA, IgD, or IgE of the same specificity as that of the initially secreted IgM. In this switch, the light chain and the variable region of the heavy chain are unchanged. Only the constant region of the heavy chain changes. This step in the differentiation of an antibody-producing cell is called *class switching* (Figure 34.23). In undifferentiated cells, the genes for the constant region of each class of heavy chain, called C_μ, C_δ, C_γ, C_ε, and C_α, are next to one another. There are eight in all, including four genes for the constant regions of γ chains. A complete transcript for the heavy chain of an IgM antibody is formed by the translocation of a V_H gene segment to a DJ_H gene segment within the genomic DNA (Figure 34.20), followed by RNA splicing of the V_HDJ_H segment to the C_μ segment, analogous to what is shown in Figure 34.19.

How are other heavy chains formed? Class switching is mediated by a gene-rearrangement process that moves a VDJ gene from a site near one C gene to a site near another C gene. Importantly, *the antigen-binding specificity is conserved in class switching because the entire V_HDJ_H gene is translocated in an intact form*. For example, the antigen-combining specificity of IgA produced by a particular cell is the same as that of IgM synthesized at an earlier stage of its development. The biological significance of C_H switching is that a whole recognition domain (the variable domain) is shifted from the early constant region (C_μ) to one of several other constant regions that mediate different effector functions.

34.4 Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors

Soluble antibodies are highly effective against extracellular pathogens, but they confer little protection against microorganisms that are predominantly intracellular, such as some viruses and mycobacteria (which cause tuberculosis and leprosy). These pathogens are shielded from antibodies by the host-cell membrane (Figure 34.24). A different and more subtle strategy, *cell-mediated immunity*, evolved to cope with intracellular pathogens. *T cells* continually scan the surfaces of all cells and kill those that exhibit foreign markings. The task is not simple; intracellular microorganisms are not so obliging as to intentionally leave telltale traces on the surface of their host. Quite the contrary, successful pathogens are masters of the art of camouflage. Vertebrates have evolved an ingenious mechanism—cut and display—to reveal the presence of stealthy intruders. Nearly all vertebrate cells exhibit on their surfaces a sample of peptides derived from the digestion of proteins in their cytoplasm. These peptides are

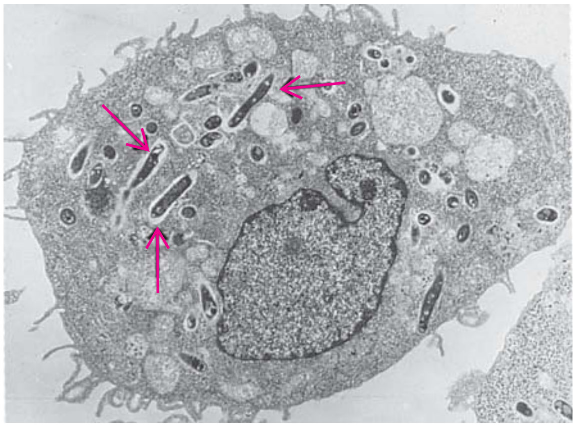


FIGURE 34.24 Intracellular pathogen. An electron micrograph showing mycobacteria (arrows) inside an infected macrophage. [Courtesy of Dr. Stanley Falkow.]

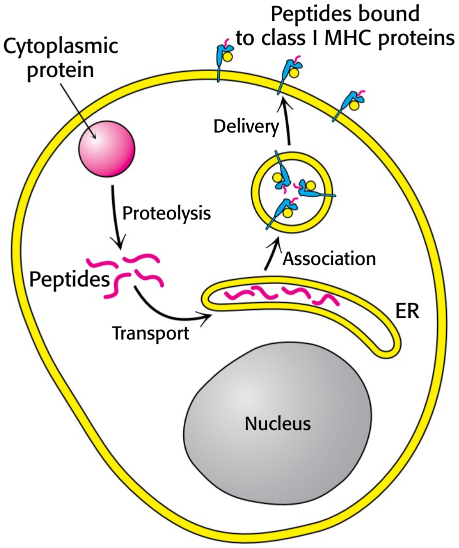


FIGURE 34.25 Presentation of peptides from cytoplasmic proteins. Class I MHC proteins on the surfaces of most cells display peptides that are derived from cytoplasmic proteins by proteolysis.

displayed by integral membrane proteins that are encoded by the *major histocompatibility complex* (MHC). Specifically, peptides derived from cytoplasmic proteins are bound to and displayed by *class I MHC proteins*. The dendritic cells of the innate immune system that subject pathogens to phagocytosis migrate to lymphatic tissue, where they use an MHC-like mechanism to present foreign peptides or lipid components to T cells—thus linking the innate and adaptive immune responses to pathogens.

How are these peptides generated and delivered to the plasma membrane? The process starts in the cytoplasm with the degradation of proteins—self-proteins as well as those of pathogens (Figure 34.25). Digestion is carried out by proteasomes (Section 23.2). The resulting peptide fragments are transported from the cytoplasm into the lumen of the endoplasmic reticulum by the *TAP protein* (for *t*ransporter associated with *a*ntigen *p*rocessing), a member of the ABC transporter family of ATP-driven pumps (Section 13.2). In the ER, peptides combine with nascent class I MHC proteins; these complexes are then targeted to the plasma membrane.

MHC proteins embedded in the plasma membrane tenaciously grip their bound peptides so that they can be touched and scrutinized by T-cell receptors on the surface of a killer cell. Foreign peptides bound to class I MHC proteins signal that a cell is infected and mark it for destruction by cytotoxic T cells. An assembly consisting of the foreign peptide—MHC complex, the T-cell receptor, and numerous accessory proteins—triggers a cascade that induces apoptosis in the infected cell. Strictly speaking, infected cells are not killed but, instead, are triggered to commit suicide to benefit the host.

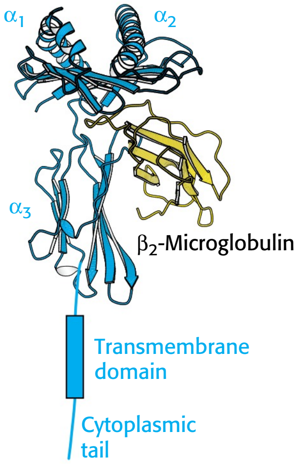


FIGURE 34.26 Class I MHC protein. A protein of this class consists of two chains. *Notice* that the α chain begins with two domains (α_1, α_2) that include α helices and continues with an immunoglobulin domain (α_3), a transmembrane domain, and a cytoplasmic tail. The second chain, β_2-microglobulin, adopts an immunoglobulin fold. [Drawn from 1HHK.pdb.]

Peptides presented by MHC proteins occupy a deep groove flanked by alpha helices

The three-dimensional structure of a large extracellular fragment of a human MHC class I protein, *human leukocyte antigen A2* (HLA-A2), was solved in 1987 by Don Wiley and Pamela Bjorkman. Class I MHC proteins consist of a 44-kDa α chain noncovalently bound to a 12-kDa polypeptide called *β_2-microglobulin*. The α chain has three extracellular domains (α_1, α_2, and α_3), a transmembrane segment, and a tail that extends into the cytoplasm (Figure 34.26). The β_2-microglobulin and the α_3 domains have immunoglobulin folds, although the pairing of the two domains differs

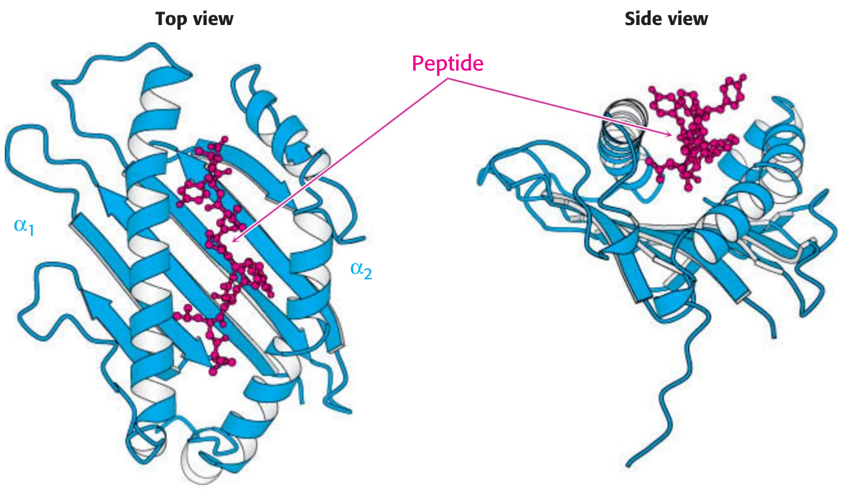


FIGURE 34.27 Class I MHC peptide-binding site. The α_1 and α_2 domains come together to form a groove in which peptides are displayed. *Notice* that that the peptide is surrounded on three sides by a β sheet and two α helices, but it is accessible from the top of the structure. [Drawn from 1HHK.pdb.]

from that in antibodies. The α_1 and α_2 domains exhibit a novel and remarkable architecture. They associate intimately to form a deep groove that serves as the peptide-binding site (Figure 34.27). The floor of the groove, which is about 25 Å long and 10 Å wide, is formed by eight β strands, four from each domain. A long helix contributed by the α_1 domain forms one side, and a helix contributed by the α_2 domain forms the other side. *This groove is the binding site for the presentation of peptides.*

The groove can be filled by a peptide from 8 to 10 residues long in an extended conformation. As we shall see (p. 1002), MHC proteins are remarkably diverse in the human population; each person expresses as many as six distinct class I MHC proteins, and many different forms are present in different people. The first structure determined, HLA-A2, binds peptides that almost always have leucine in the second position and valine in the last position (Figure 34.28). Side chains from the MHC molecule interact with the amino and carboxyl termini and with the side chains in these two key positions. These two residues are often referred to as the *anchor residues*. The other residues are highly variable. Thus, many millions of different peptides can be presented by this particular class I MHC protein; the

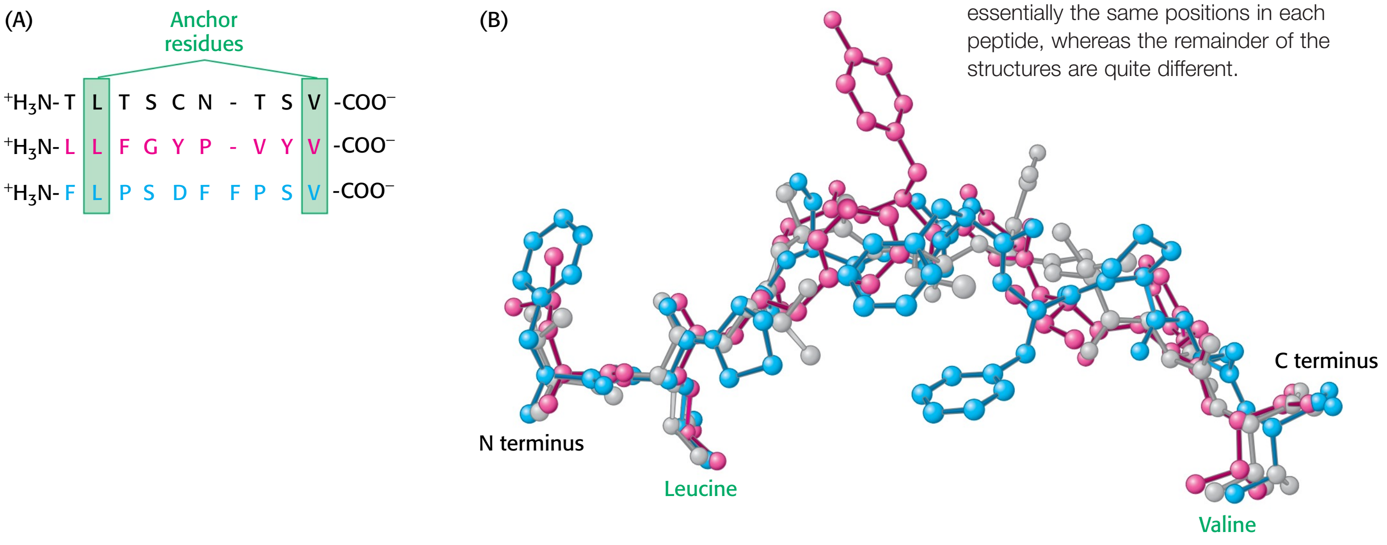

(A)
Anchor residues
^+H_3N- T L T S C N - T S V -COO$^-$
^+H_3N- L L F G Y P - V Y V -COO$^-$
^+H_3N- F L P S D F F P S V -COO$^-$
(B)
N terminus
Leucine
C terminus
Valine

FIGURE 34.28 Anchor residues. (A) The amino acid sequences of three peptides that bind to the class I MHC protein HLA-A2 are shown. Each of these peptides has leucine in the second position and valine in the carboxyl-terminal position. (B) Comparison of the structures of these peptides reveals that the amino and carboxyl termini, as well as the side chains of the leucine and valine residues, are in essentially the same positions in each peptide, whereas the remainder of the structures are quite different.

identities of only two of the nine residues are crucial for binding. Each class of MHC molecules requires a unique set of anchor residues. Thus, a tremendous range of peptides can be presented by these molecules. Note that *one face of the bound peptide is exposed to solution, where it can be examined by other molecules, particularly T-cell receptors*. An additional remarkable feature of MHC–peptide complexes is their kinetic stability; once bound, a peptide is not released, even after a period of days.

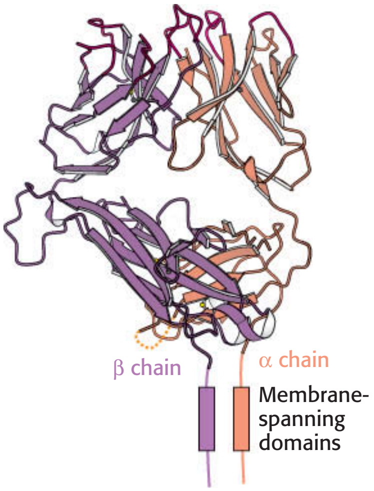


FIGURE 34.29 T-cell receptor. This protein consists of an α chain and a β chain, linked by a disulfide bond. *Notice* that each chain consists of two immunoglobulin domains on the cell surface, a membrane-spanning domain, and a short cytoplasmic tail. [Drawn from 1BD2.pdb.]

T-cell receptors are antibody-like proteins containing variable and constant regions

We are now ready to consider the receptor that recognizes peptides displayed by MHC proteins on target cells. The *T-cell receptor* consists of a 43-kDa α chain joined by a disulfide bond to a 43-kDa β chain (Figure 34.29). Each chain spans the plasma membrane and has a short carboxyl-terminal region on the cytoplasmic side. A small proportion of T cells express a receptor consisting of γ and δ chains in place of α and β. The α and β chains of the T-cell receptor, like immunoglobulin L and H chains, consist of *variable* and *constant* regions. Indeed, *these domains of the T-cell receptor are homologous to the V and C domains of immunoglobulins*. Furthermore, hypervariable sequences present in the V regions of the α and β chains of the T-cell receptor form the binding site for the epitope.

The genetic architecture of these proteins is similar to that of immunoglobulins, though the antibody genetic diversity is distributed over all the CDR loops, whereas T-cell-receptor genetic diversity is concentrated in the CDR3 loop that interacts with the peptide bound to the MHC. The variable region of the T-cell receptor α chain is encoded by about 50 V gene segments and 70 J gene segments. The T-cell receptor β chain is encoded by two D gene segments in addition to 57 V gene segments and 13 J gene segments. Again, the diversity of gene segments and the use of slightly imprecise modes of joining them increase the number of distinct proteins formed. *At least 10^{12} different specificities could arise from combinations of this repertoire of genes*. Thus, T-cell receptors, like immunoglobulins, can recognize a very large number of different epitopes. All the receptors on a particular T cell have the same specificity.

How do T cells recognize their targets? The variable regions of the α and β chains of the T-cell receptor form a binding site that recognizes a combined epitope–foreign peptide bound to an MHC protein (Figure 34.30). Neither the foreign peptide alone nor the MHC protein alone forms a complex with the T-cell receptor. Thus, fragments of an intracellular pathogen are presented in a context that allows their detection, leading to the initiation of an appropriate response.

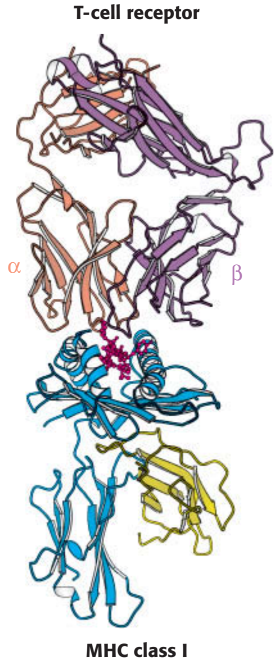


FIGURE 34.30 T-cell receptor–class I MHC complex. The T-cell receptor binds to a class I MHC protein containing a bound peptide (red). *Notice* that the T-cell receptor contacts both the MHC protein and the peptide. [Drawn from 1BD2.pdb.]

CD8 on cytotoxic T cells acts in concert with T-cell receptors

The T-cell receptor does not act alone in recognizing and mediating the fate of target cells. Cytotoxic T cells also express a protein termed *CD8* on their surfaces that is crucial for the recognition of the class I MHC–peptide complex. The abbreviation CD stands for *cluster of differentiation*, referring to a cell-surface marker that is used to identify a lineage or stage of differentiation. Antibodies specific for particular CD proteins have been invaluable in following the development of leukocytes and in discovering new interactions between specific cell types.

Each chain in the CD8 dimer contains a domain that resembles an immunoglobulin variable domain (Figure 34.31). CD8 interacts primarily with the constant α_3 domain of class I MHC proteins. This interaction further stabilizes the interactions between the T cell and its target.

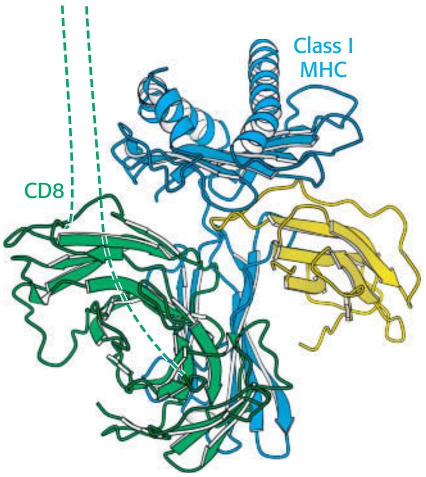


FIGURE 34.31 The coreceptor CD8. This dimeric protein extends from the surface of a cytotoxic T cell and binds to class I MHC molecules that are expressed on the surface of the cell that is bound to the T cell. The dashed lines represent extended polypeptide chains that link the immunoglobulin domains of CD8 to the membrane. *Notice* that the coreceptor interacts primarily with the constant α_3 domain of the class I MHC domain. [Drawn from 1AKJ.pdb.]

The cytoplasmic tail of CD8 contains a docking site for Lck, a cytoplasmic tyrosine kinase akin to Src. The T-cell receptor itself is associated with six polypeptides that form the CD3 complex (Figure 34.32). The γ, δ, and ε chains of CD3 are homologous to Ig-α and Ig-β associated with the B-cell receptor (Figure 34.21); each chain consists of an extracellular immunoglobulin domain and an intracellular ITAM region. These chains associate into CD3-γε and CD3-δε heterodimers. An additional component, the CD3-ζ chain, has only a small extracellular domain and a larger intracellular domain containing three ITAM sequences.

On the basis of these components, a model for T-cell activation can be envisaged that is closely parallel to the pathway for B-cell activation (Section 34.3; Figure 34.33). The binding of the T-cell receptor to the class I MHC–peptide complex and the concomitant binding of CD8 from the T cell to the MHC molecule link the kinase Lck to the ITAM substrates of the components of the CD3 complex. The phosphorylation of the tyrosine residues in the ITAM sequences generates docking sites for a protein kinase called ZAP-70 (for 70-kDa *zeta-associated protein*) that is homologous to Syk in B cells. Docked by its two SH2 domains, ZAP-70 phosphorylates

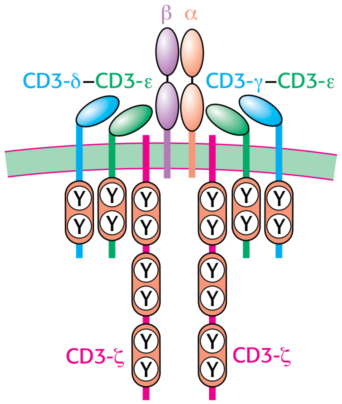


FIGURE 34.32 T-cell receptor complex. The T-cell receptor is associated with six CD3 molecules: a CD3-γ–CD3-ε heterodimer, a CD3-δ–CD3-ε heterodimer, and two chains of CD3-ζ. Single ITAM sequences are present in the cytoplasmic domains of CD3-γ, CD3-δ, and CD3-ε, whereas three such sequences are found in each CD3-ζ chain.

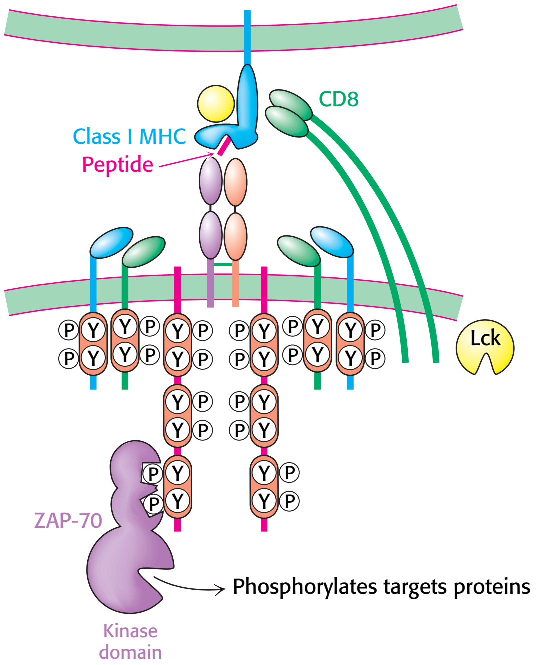


FIGURE 34.33 T-cell activation. The interaction between the T-cell receptor and a class I MHC–peptide complex results in the binding of CD8 to the MHC protein, the recruitment of the protein tyrosine kinase Lck, and the phosphorylation of tyrosine residues in the ITAM sequences of the CD3 chains. After phosphorylation, the ITAM regions serve as docking sites for the protein kinase ZAP-70, which phosphorylates protein targets to transmit the signal.

downstream targets in the signaling cascade. Additional molecules, including a membrane-bound protein phosphatase called CD45 and a cell-surface protein called CD28, play ancillary roles in this process.

T-cell activation has two important consequences. First, the activation of cytotoxic T cells results in the secretion of several proteins, including *perforin* and *granzymes*. Perforin is a 70-kDa protein that destabilizes the plasma membrane of the target cell, enabling the entry of *granzymes* into the cytoplasm of the target cell. Granzymes are serine proteases (Section 9.1) that initiate the pathway of apoptosis, leading to the death of the target cell and the fragmentation of its DNA, including any viral DNA that may be present. Second, after it has stimulated its target cell to commit suicide, the activated T cell disengages and is stimulated to reproduce. Thus, additional T cells that express the same T-cell receptor are generated to continue the battle against the invader after these T cells have been identified as a suitable weapon.

Helper T cells stimulate cells that display foreign peptides bound to class II MHC proteins

Not all T cells are cytotoxic. *Helper T cells* stimulate the proliferation of specific B lymphocytes and cytotoxic T cells and thereby serve as partners in determining the immune responses that are produced. The importance of helper T cells is graphically revealed by the devastation wrought by AIDS, a condition that destroys these cells. Helper T cells, like cytotoxic T cells, detect foreign peptides that are presented on cell surfaces by MHC proteins. However, the source of the peptides, the MHC proteins that bind them, and the transport pathway are different.

Helper T cells recognize peptides bound to MHC molecules referred to as class II. Their helping action is focused on B cells, macrophages, and dendritic cells. *Class II MHC proteins are expressed only by these antigen-presenting cells*, unlike class I MHC proteins, which are expressed on nearly all cells. The peptides presented by class II MHC proteins do not come from the cytoplasm. Rather, *they arise from the degradation of proteins that have been internalized by endocytosis*. Consider, for example, a virus particle that is captured by membrane-bound immunoglobulins on the surface of a B cell (Figure 34.34). This complex is delivered to an endosome, a membrane-enclosed acidic compartment, where it is digested. The resulting peptides become associated with class II MHC proteins, which move to the cell surface. Peptides from the cytoplasm cannot reach class II proteins, whereas peptides from endosomal compartments cannot reach class I proteins. This segregation of displayed peptides is biologically critical. The association of a foreign peptide with a class II MHC protein signals that a cell has *encountered* a pathogen and serves as a call for *help*. In contrast, association with a class I MHC protein signals that a cell has *succumbed* to a pathogen and is a call for *destruction*.

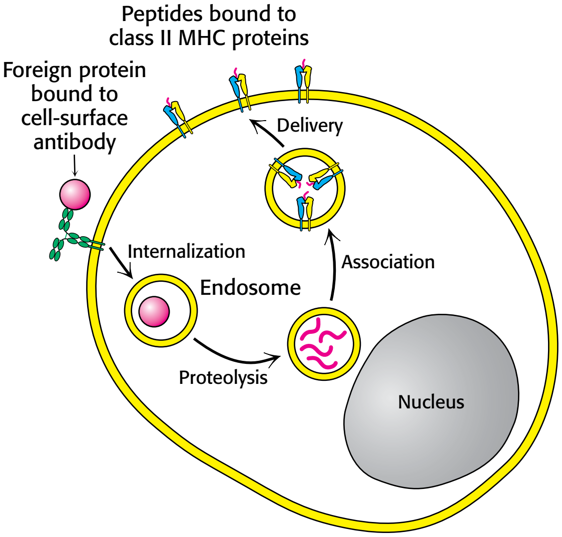

Peptides bound to class II MHC proteins
Foreign protein bound to cell-surface antibody
Delivery
Internalization
Endosome
Association
Proteolysis
Nucleus

FIGURE 34.34 Presentation of peptides from internalized proteins. Antigen-presenting cells bind and internalize foreign proteins and display peptides that are formed from the digestion of these proteins in class II MHC proteins.

Helper T cells rely on the T-cell receptor and CD4 to recognize foreign peptides on antigen-presenting cells

The overall structure of a class II MHC molecule is remarkably similar to that of a class I molecule. Class II molecules consist of a 33-kDa α chain and a noncovalently bound 30-kDa β chain (Figure 34.35). Each contains two extracellular domains, a transmembrane segment, and a short cytoplasmic tail. The peptide-binding site is formed by the α_1 and β_1 domains, each of

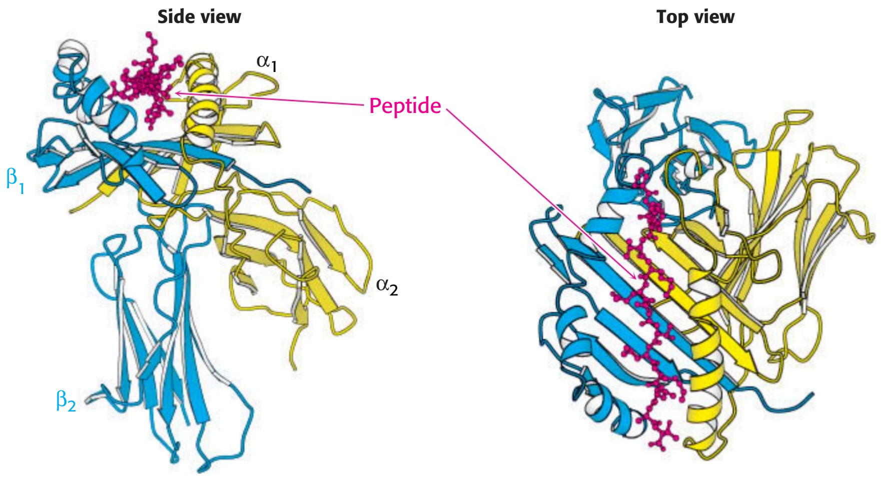


FIGURE 34.35 Class II MHC protein. A class II MHC protein consists of homologous α and β chains, each of which has an amino-terminal domain that constitutes half of the peptide-binding structure, as well as a carboxyl-terminal immunoglobulin domain. *Notice* the trough-like peptide-binding site, which is similar to that in class I MHC proteins except that it is open at both ends, allowing class II MHC proteins to bind longer peptides than those bound by class I. [Drawn from 1DLH.pdb.]

which contributes a long helix and part of a β sheet. Thus, the same structural elements are present in class I and class II MHC molecules, but they are combined into polypeptide chains in different ways. The peptide-binding site of a class II molecule is open at both ends, and so this groove can accommodate longer peptides than can be bound by class I molecules; typically, peptides between 13 and 18 residues long are bound. The peptide-binding specificity of each class II molecule depends on binding pockets that recognize particular amino acids, also known as anchor residues, in specific positions along the sequence.

Helper T cells express T-cell receptors that are produced from the same genes as those on cytotoxic T cells. These T-cell receptors interact with class II MHC molecules in a manner that is analogous to T-cell-receptor interaction with class I MHC molecules. Nonetheless, helper T cells and cytotoxic T cells are distinguished by other proteins that they express on their surfaces. In particular, helper T cells express a protein called *CD4* instead of expressing CD8. CD4 consists of four immunoglobulin domains that extend from the T-cell surface, as well as a small cytoplasmic region (Figure 34.36). The amino-terminal immunoglobulin domains of CD4 interact with the base on the class II MHC molecule. Thus, helper T cells bind cells expressing class II MHC specifically because of the interactions with CD4 (Figure 34.37).

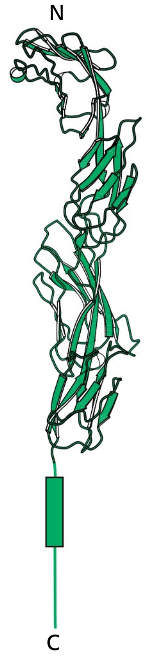


FIGURE 34.36 Coreceptor CD4. This protein comprises four tandem immunoglobulin domains that extend from the surface of a helper T cell. [Drawn from 1WIO.pdb.]

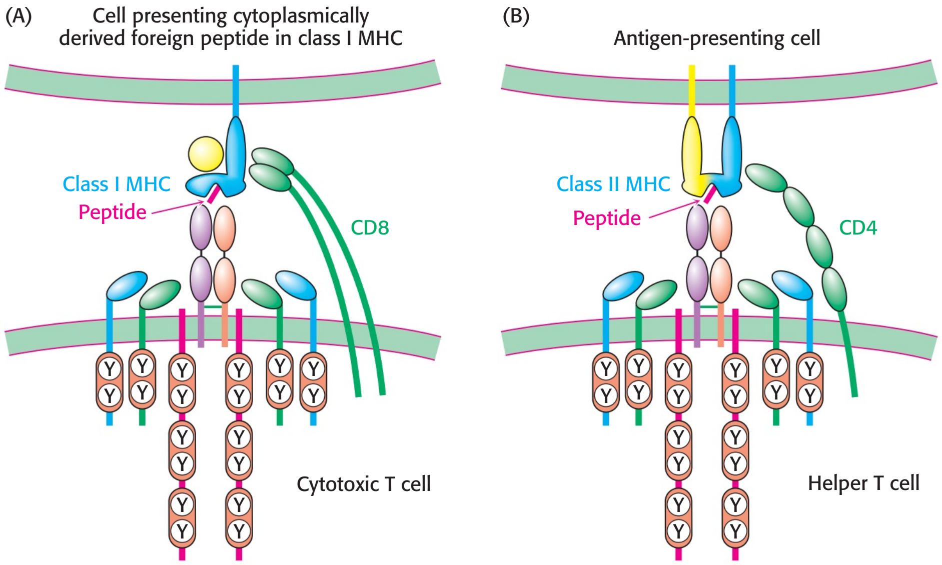


FIGURE 34.37 Variations on a theme. (A) Cytotoxic T cells recognize foreign peptides presented in class I MHC proteins with the aid of the coreceptor CD8. (B) Helper T cells recognize peptides presented in class II MHC proteins by specialized antigen-presenting cells with the aid of the coreceptor CD4.

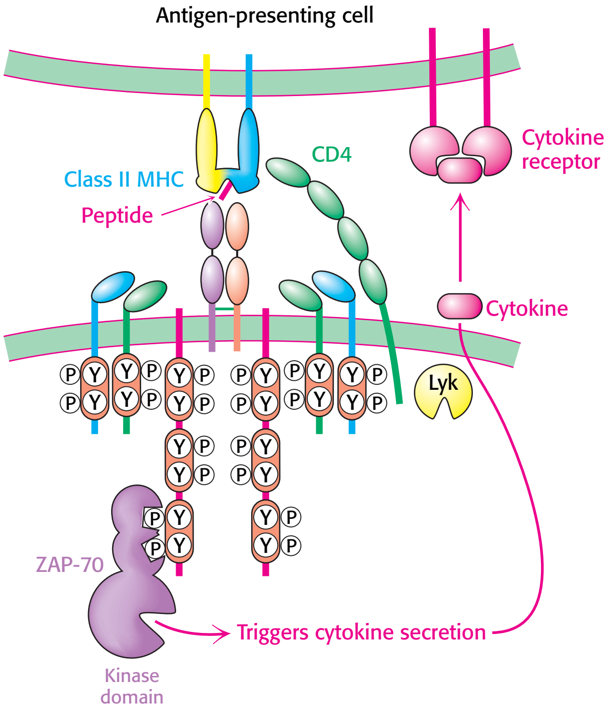


FIGURE 34.38 Helper-T-cell action. The engagement of the T-cell receptor in helper T cells results in the secretion of cytokines. These cytokines bind to cytokine receptors expressed on the surface of the antigen-presenting cell, stimulating cell growth, differentiation, and, in regard to a B cell, antibody secretion.

When a helper T cell binds to an antigen-presenting cell expressing an appropriate class II MHC–peptide complex, signaling pathways analogous to those in cytotoxic T cells are initiated by the action of the kinase Lck on ITAMs in the CD3 molecules associated with the T-cell receptor. However, rather than triggering events leading to the death of the attached cell, *these signaling pathways result in the secretion of cytokines from the helper cell.* Cytokines are a family of molecules that include, among others, interleukin-2 and interferon-γ. Cytokines bind to specific receptors on the antigen-presenting cell and stimulate its growth and differentiation (Figure 34.38). For example, these cytokines stimulate the differentiation of B cells into *plasma cells*, which specialize in the production of large amounts of antibody. Thus, the internalization and presentation of parts of a foreign pathogen help to generate a local environment in which cells taking part in the defense against this pathogen can flourish through the action of helper T cells.

MHC proteins are highly diverse

MHC class I and II proteins, the presenters of peptides to T cells, were discovered because of their role in *transplantation rejection.* A tissue transplanted from one person to another or from one mouse to another is usually rejected by the immune system. In contrast, tissues transplanted from one identical twin to another or between mice of an inbred strain are accepted. Genetic analyses revealed that rejection occurs when tissues are transplanted between individual organisms having different genes in the major histocompatibility complex, a cluster of more than 75 genes playing key roles in immunity. The 3500-kb span of the MHC is nearly the length of the entire *E. coli* chromosome. The MHC encodes class I proteins (presenters to cytotoxic T cells) and class II proteins (presenters to helper T cells), as well as class III proteins (components of the complement cascade) and many other proteins that play key roles in immunity.

Human beings express six different class I genes (three from each parent) and six different class II genes. The three loci for class I genes are called HLA-A, -B, and -C; those for class II genes are called HLA-DP, -DQ, and -DR. These loci are *highly polymorphic:* many alleles of each are present in the population. For example, more than 50 each of HLA-A, -B, and -C alleles are known; the numbers discovered increase each year. Hence, the likelihood that two unrelated persons have identical class I and II proteins is very small ($<10^{-4}$), accounting for transplantation rejection unless the genotypes of donor and acceptor are closely matched in advance.

Differences between class I proteins are located mainly in the α_1 and α_2 domains, which form the peptide-binding site (Figure 34.39). The α_3 domain, which interacts with a constant β_2-microglobulin, is largely conserved. Similarly, the differences between class II proteins cluster near the peptide-binding groove. Why are MHC proteins so highly variable? *Their diversity makes the presentation of a very wide range of peptides to T cells possible*. A *particular* class I or class II molecule may not be able to bind any of the peptide fragments of a viral protein. The likelihood of a fit is markedly increased by having several kinds (usually six) of each class of presenters in each individual organism. If all members of a species had identical class I or class II molecules, the population would be much more vulnerable to devastation by a pathogen that had mutated and thereby evaded presentation. The evolution of the diverse human MHC repertoire has been driven by the selection for individual members of the species who resist infections to which other members of the population may be susceptible.

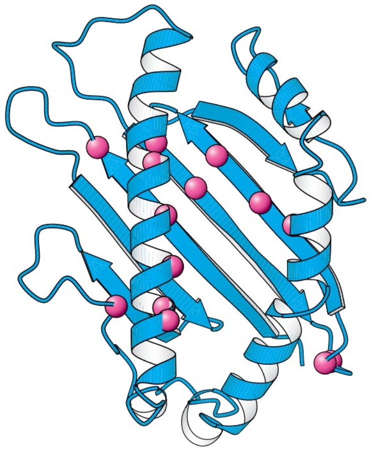


FIGURE 34.39 Polymorphism in class I MHC proteins. *Notice* that the positions of sites with a high degree of polymorphism in the human population are displayed as red spheres on the structure of the amino-terminal part of a class I MHC protein. [Drawn from 1HHK.pdb.]

Human immunodeficiency viruses subvert the immune system by destroying helper T cells

In 1981, the first cases of a new disease now called *acquired immune deficiency syndrome* (AIDS) were recognized. The victims died of rare infections because their immune systems were crippled. The cause was identified 2 years later by Luc Montagnier and coworkers. AIDS is produced by *human immunodeficiency virus* (HIV), of which two major classes are known: HIV-1 and the much less common HIV-2. Like other *retroviruses,* HIV contains a single-stranded RNA genome that is replicated through a double-stranded DNA intermediate. This viral DNA becomes integrated into the genome of the host cell. In fact, viral genes are transcribed only when they are integrated into the host DNA.

The HIV virion is enveloped by a lipid-bilayer membrane containing two glycoproteins: gp41 spans the membrane and is associated with gp120, which is located on the external face (Figure 34.40). The core of the virus contains two copies of the RNA genome and associated transfer RNAs, as well as several molecules of reverse transcriptase. They are surrounded by many copies of two proteins called p17 and p24. *The host cell for HIV is the helper T cell*. The gp120 molecules on the membrane of HIV bind to CD4 molecules on the surface of the helper T cell (Figure 34.41). This interaction allows the associated viral gp41 to insert its amino-terminal head into the host-cell membrane. The viral membrane and the helper-T-cell membrane fuse, and the viral core is released directly into the cytoplasm. Infection by

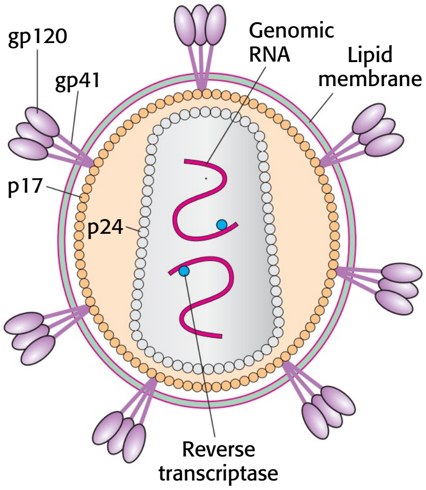


FIGURE 34.40 Human immunodeficiency virus. A schematic representation of HIV reveals its proteins and nucleic acid components. [Information from G. B. Karlsson Hedestam et al., *Nat. Rev. Microbiol.* 6:143–155, 2008, Fig. 2a.]

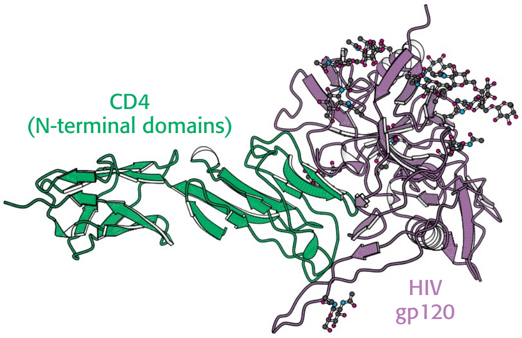


FIGURE 34.41 HIV receptor. A complex between a modified form of the envelope glycoprotein gp120 from HIV and a peptide corresponding to the two amino-terminal domains from the helper-T-cell protein CD4 reveals how viral infection of helper T cells is initiated. [Drawn from 1GC1.pdb.]

HIV leads to the destruction of helper T cells because the permeability of the host plasma membrane is markedly increased by the insertion of viral glycoproteins and the budding of virus particles. The influx of ions and water disrupts the ionic balance, causing osmotic lysis.

34.5 The Immune System Contributes to the Prevention and the Development of Human Diseases

The primary function of the immune system is to protect the host from invasion by foreign organisms. But how does the immune system prevent itself from mounting attacks against its host? In other words, how does the immune system distinguish between self and nonself? Clearly, proteins from the organism itself do not bear some special tag identifying them. Instead, selection processes early in the developmental pathways for immune cells kill or suppress those immune cells that react strongly with self-antigens. The evolutionary paradigm still applies; immune cells that recognize self-antigens are generated, but selective mechanisms eliminate such cells in the course of development.

T cells are subjected to positive and negative selection in the thymus

T cells derive their name from the location of their production—the thymus, a small organ situated just above the heart. Examination of the developmental pathways leading to the production of mature cytotoxic and helper T cells reveals the selection mechanisms that are crucial for distinguishing self from nonself. These selection criteria are quite stringent; approximately 98% of the thymocytes, the precursors of T cells, die before the completion of the maturation process.

Thymocytes produced in the bone marrow do not express the T-cell receptor complex, CD4, or CD8. On relocation to the thymus and rearrangement of the T-cell-receptor genes, the immature thymocyte expresses all of these molecules. These cells are first subjected to *positive selection* (Figure 34.42). Cells for which the T-cell receptor can bind with reasonable affinity to either class I or class II MHC molecules survive this selection; those for which the T-cell receptor does not participate in such an interaction undergo apoptosis and die. *The role of the positive selection step is to prevent the production of T cells that will not bind to any MHC complex present, regardless of the peptide bound.*

The cell population that survives positive selection is subjected to a second step, *negative selection.* Here, T cells that bind with high affinity to MHC complexes bound to self-peptides expressed on the surfaces of

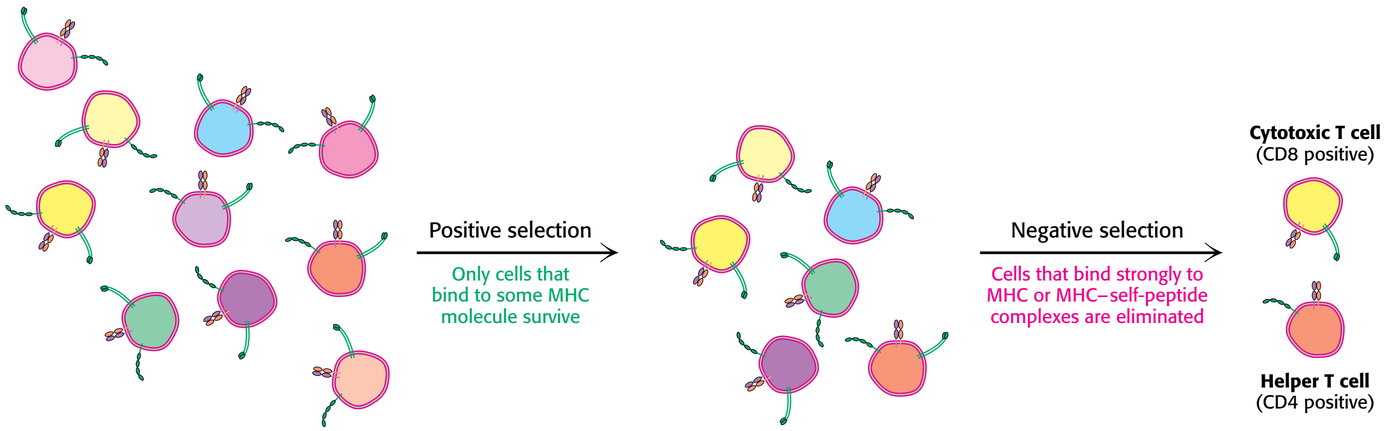

Positive selection
Only cells that bind to some MHC molecule survive
Negative selection
Cells that bind strongly to MHC or MHC–self-peptide complexes are eliminated
Cytotoxic T cell
(CD8 positive)
Helper T cell
(CD4 positive)

FIGURE 34.42 T cell selection. A population of thymocytes is subjected first to positive selection to remove cells that express T-cell receptors that will not bind to MHC proteins expressed by the individual organism. The surviving cells are then subjected to negative selection to remove cells that bind strongly to MHC complexes bound to self-peptides.

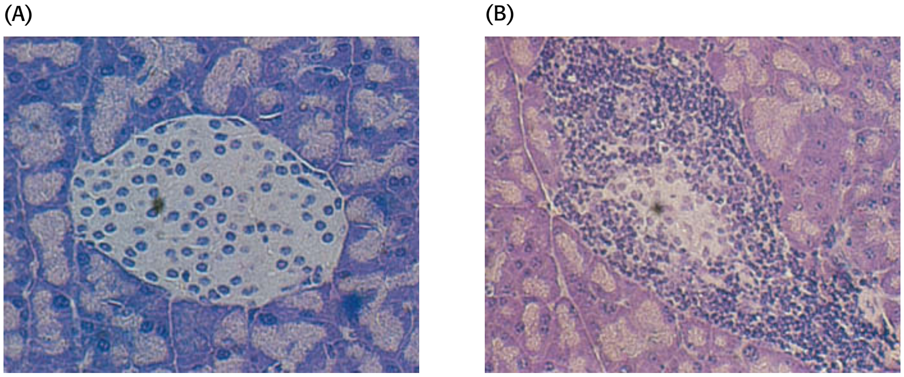


FIGURE 34.43 Consequences of autoimmunity. Photomicrographs of an islet of Langerhans (A) in the pancreas of a normal mouse and (B) in the pancreas of a mouse with an immune response against pancreatic β cells, which results in a disease resembling insulin-dependent diabetes mellitus in human beings. *Notice* that the relatively pale cellular area in the center of part A is populated with the dark nuclei of inflammatory cells in part B. [From M. A. Atkinson and N. K. Maclaren. What causes diabetes? Scientific American, 1990. Courtesy Mark Atkinson.]

antigen-presenting cells in the thymus undergo apoptosis or are otherwise suppressed. Those that do not bind too avidly to any such MHC complex complete development and become mature cytotoxic T cells (which express only CD8) or helper T cells (which express only CD4). The negative selection step leads to *self-tolerance;* cells that bind an MHC–self-peptide complex are removed from the T-cell population. Similar mechanisms apply to developing B cells: maturing B cells that express antibodies that interact strongly with self-antigens are suppressed.

Autoimmune diseases result from the generation of immune responses against self-antigens

Although thymic selection is remarkably efficient in suppressing the immune response to self-antigens, failures do occur. Such failures result in *autoimmune diseases.* These diseases include common illnesses such as insulin-dependent diabetes mellitus, multiple sclerosis, and rheumatoid arthritis. In these illnesses, immune responses against self-antigens result in damage to selective tissues that express the antigen (Figure 34.43).

In many cases, the cause of the generation of self-reactive antibodies or T cells is unclear. However, in other cases, infectious organisms such as bacteria or viruses may play a role. Infection leads to the generation of antibodies and T cells that react with many different epitopes from the infectious organism. If one of these antigens closely resembles a self-antigen, an autoimmune response can result. For example, *Streptococcus* infections sometimes lead to rheumatic fever owing to the production of antibodies to streptococcal antigens that cross-react with exposed epitopes in heart muscle.

The immune system plays a role in cancer prevention

The development of immune responses against proteins encoded by our own genomes can be beneficial in some circumstances. Cancer cells have undergone significant changes that often result in the production of proteins that are not normally expressed. For example, the mutation of genes can generate proteins that do not correspond in amino acid sequence to any normal protein. Such proteins may be recognized as foreign, and an immune response will be generated specifically against the cancer cell. Alternatively, cancer cells often produce proteins that are expressed during embryonic development but are not expressed or are expressed at very low levels after birth. For example, a membrane glycoprotein called *carcinoembryonic antigen* (CEA) appears in the gastrointestinal cells of developing fetuses but is not normally expressed at significant levels after birth. More than 50% of patients with colorectal cancer have elevated serum levels of CEA. Immune cells recognizing epitopes from such proteins will not be subjected to negative selection and, hence, will be present in the adult immune repertoire. These cells may play a cancer surveillance role, killing cells that overexpress antigens such as CEA and preventing genetically damaged cells from developing into tumors.

Vaccines are a powerful means to prevent and eradicate disease

The discovery and utilization of *vaccines*, biological preparations that stimulate immunity to a particular pathogen or disease, represents a significant milestone in modern medical history. The pioneering work of Edward Jenner and Louis Pasteur in the eighteenth and nineteenth centuries revealed that inoculation with inactivated forms of a pathogen could protect against subsequent infection with the active pathogen. Treatment of a person with a vaccine, or *immunization*, leads to the development of *immunological memory*, the ability of the immune system to respond more rapidly and effectively to pathogens that have been encountered previously. At the cellular level, memory arises because positively selected B cells and T cells that have a high affinity for their target persist within the organism as *memory B cells* and *memory T cells*.

In order for a vaccine to be effective, it must generate a sustained protective response appropriate for the targeted pathogen. For example, pathogens such as *Clostridium tetani*, the causative agent of tetanus, cause disease by the release of destructive proteins called *toxins*. Hence, an effective tetanus vaccine should generate a strong B-cell-mediated response. In contrast, a vaccine designed to prevent disease caused by an intracellular pathogen should generate a strong T-cell-mediated response. The active component of a vaccine that is responsible for stimulating this immune response can take many forms:

Killed or *inactivated vaccines* contain pathogens that have been rendered harmless by treatment with chemicals or high heat.

Live attenuated vaccines contain live pathogens that have accumulated mutations so that they are no longer virulent to human cells. These vaccines are most commonly generated by repeated infection of the pathogen in cultured cells until its virulence is lost.

Subunit vaccines contain a purified protein component of the pathogen. Such proteins can be either isolated from infected material (such as blood from chronically infected patients) or generated by recombinant methods.

Toxoid vaccines are used against pathogens that employ an extracellular toxin to cause disease. These vaccines contain a form of the toxin that has been inactivated by treatment with chemicals or high heat.

The effectiveness of immunization efforts has made the global eradication of certain diseases possible. In 1966, 20 million cases of smallpox were reported worldwide. Just 14 years later, after coordinated vaccination efforts in regions of endemic disease, the World Health Assembly declared that smallpox had been eradicated. Nevertheless, despite the success of vaccines in the prevention of many devastating diseases, several pathogens have posed a significant challenge to vaccine development. For example, the development of an effective HIV vaccine has been complicated by the antigenic diversity of HIV strains. Because its mechanism for replication is prone to error, a population of HIV presents an ever-changing array of coat proteins. Indeed, the mutation rate of HIV is more than 65 times higher than that of influenza virus. Nevertheless, the identification of protective antibodies in individuals resistant to HIV and advancements in vaccine design suggest that the development of an effective HIV vaccine remains a real possibility.

SUMMARY

Two lines of defense against pathogens are the innate immune system and the adaptive immune system. The innate immune system targets features common to many different pathogens but misses those pathogens lacking these features. The adaptive immune system is both more specific and wide-reaching. To respond effectively to a vast array of pathogens, this type of immune system must be tremendously flexible. The adaptive immune system follows the principles of evolution: an enormously diverse set of potentially useful proteins is generated; these proteins are then subjected to intense selection so that only cells that express useful proteins flourish and continue development, until an effective immune response to a specific invader is generated.

34.1 Antibodies Possess Distinct Antigen-Binding and Effector Units

The major immunoglobulin in the serum is immunoglobulin G. An IgG protein is a heterotetramer with two heavy chains and two light chains. Treatment of IgG molecules with proteases such as papain produces three fragments: two F_{ab} fragments that retain antigen-binding activity and an F_c fragment that retains the ability to activate effector functions such as the initiation of the complement cascade. The F_{ab} fragments include the L chain and the amino-terminal half of the H chain; the F_c domain is a dimer consisting of the carboxyl-terminal halves of two H chains. Five different classes of antibody—IgG, IgM, IgA, IgD, and IgE—differ in their heavy chains and, hence, in their effector functions.

34.2 Antibodies Bind Specific Molecules Through Hypervariable Loops

One particular protein fold is found in many of the key proteins of the immune system. The immunoglobulin fold consists of a pair of β sheets that pack against each other, linked by a single disulfide bond. Loops projecting from one end of the structure form a binding surface that can be varied by changing the amino acid sequences within the loops. Domains with immunoglobulin folds are linked to form antibodies and other classes of proteins in the immune system, including T-cell receptors. Two chains come together to form the binding surface of an antibody. Three loops from each domain, the complementarity-determining regions, form an essentially continuous surface that can vary tremendously in shape, charge, and other characteristics to allow particular antibodies to bind to molecules ranging from small molecules to large protein surfaces.

34.3 Diversity Is Generated by Gene Rearrangements

The tremendous diversity of the amino acid sequences of antibodies is generated by segmental rearrangements of genes. For antibody κ light chains, 1 of 40 variable regions is linked to 1 of 5 joining regions. The combined VJ unit is then linked to the constant region. Thousands of different genes can be generated in this manner. Similar arrays are rearranged to form the genes for the heavy chains, but an additional region called the diversity region lies between the V and the J regions. The combination of L and H chains, each obtained through such rearranged genes, can produce more than 10^8 distinct antibodies. Different classes of antibodies are also generated by gene rearrangements that lead to class switching. Oligomerization of membrane-bound antibody molecules initiates a signal-transduction cascade inside B cells. Key steps in this signaling process include the phosphorylation of specific tyrosine residues in sequences termed immunoreceptor tyrosine-based activation motifs, present in proteins that associate with the membrane-bound antibodies.

34.4 Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors

Intracellular pathogens such as viruses and mycobacteria cannot be easily detected. Intracellular proteins are constantly being cut into small peptides by proteasomes and displayed in class I major-histocompatibility-complex proteins on cell surfaces. Such peptides lie in a groove defined by two helices in the class I MHC proteins. The combination of MHC protein and peptide can be bound by an appropriate T-cell receptor. T-cell receptors resemble the antigen-binding domains of antibodies in structure, and diversity in T-cell-receptor sequence is generated by V(D)J gene rearrangements. The T-cell receptor recognizes features of both the peptide and the MHC molecule that presents it.

Cytotoxic T cells initiate apoptosis in cells to which they bind through interactions between T-cell receptors and class I MHC–peptide complexes aided by interactions with the coreceptor molecule CD8. Helper T cells recognize peptides presented in class II MHC proteins, a distinct type of MHC protein expressed only on antigen-presenting cells, such as B cells and macrophages. Helper T cells express the coreceptor CD4 rather than CD8. CD4 interacts with class II MHC proteins present on antigen-presenting cells. Signaling pathways, analogous to those in B cells, are initiated by interactions between MHC–peptide complexes and T-cell receptors and the CD8 and CD4 coreceptors. Human immunodeficiency virus damages the immune system by infecting and killing cells that express CD4, such as helper T cells.

34.5 The Immune System Contributes to the Prevention and the Development of Human Diseases

In principle, the immune system is capable of generating antibodies and T-cell receptors that bind to self-molecules—that is, molecules that are normally present in a healthy and uninfected individual organism. Selection mechanisms prevent such self-directed molecules from being expressed at high levels. The selection process includes both positive selection, to enrich the population of cells that express molecules that have the potential to bind foreign antigens in an appropriate context, and negative selection, which eliminates cells that express molecules with too high an affinity for self-antigens. Autoimmune diseases such as insulin-dependent diabetes mellitus can result from the amplification of a response against a self-antigen. Vaccines stimulate immunological memory so as to prevent, and even eradicate, disease.

KEY TERMS

innate immune system (p. 981)
adaptive immune system (p. 981)
phagocyte (p. 982)
Toll-like receptor (TLR) (p. 982)
pathogen-associated molecular pattern (PAMP) (p. 982)
endotoxin (p. 982)
humoral immune response (p. 984)
B lymphocyte (B cell) (p. 984)

antigen (p. 984)
antigenic determinant (epitope) (p. 984)
cellular immune response (p. 984)
cytotoxic T lymphocyte (killer T cell) (p. 984)
helper T lymphocyte (helper T cell) (p. 984)
immunoglobulin G (IgG) (p. 985)
F_{ab} (p. 985)
F_c (p. 985)
light (L) chain (p. 985)
heavy (H) chain (p. 985)
segmental flexibility (p. 987)
immunoglobulin M (IgM) (p. 987)
immunoglobulin A (IgA) (p. 987)
immunoglobulin D (IgD) (p. 987)
immunoglobulin E (IgE) (p. 987)
variable region (p. 988)
constant region (p. 988)
immunoglobulin fold (p. 988)
hypervariable loop (p. 988)
complementarity-determining region (CDR) (p. 988)
V(D)J recombination (p. 992)
immunoreceptor tyrosine-based activation motif (ITAM) (p. 993)
cyclosporin (p. 994)
hapten (p. 994)
class switching (p. 995)
major histocompatibility complex (MHC)(p. 996)
class I MHC protein (p. 996)
TAP (transporter associated with antigen processing) protein (p. 996)
human leukocyte antigen (HLA) (p. 996)
β_2-microglobulin (p. 996)
T-cell receptor (p. 998)
CD8 (p. 998)
perforin (p. 1000)
granzyme (p. 1000)
class II MHC protein (p. 1000)
CD4 (p. 1001)
plasma cell (p. 1002)
human immunodeficiency virus (HIV) (p. 1003)
positive selection (p. 1004)
negative selection (p. 1004)
autoimmune disease (p. 1005)
carcinoembryonic antigen (CEA) (p. 1006)
vaccine (p. 1006)
immunological memory (p. 1006)
killed, or inactivated, vaccines (p. 1006)
live attenuated vaccines (p. 1006)
subunit vaccines (p. 1006)
toxoid vaccines (p. 1006)

PROBLEMS

1. *First things first.* Distinguish between the innate and adaptive immune systems.

2. *Antibody diversity.* What are the mechanisms used by B cells to generate antibody diversity?

3. *Hang in there.* Explain the difference between *affinity* and *avidity.* For which immunoglobulin class might avidity be particularly important in antigen recognition?

4. *Innate abilities.* A strain of mice has been identified that does not respond to LPS. This lack of response is due to a single amino acid change in the intracellular domain of mouse TLR4. Propose an explanation for the lack of response.

5. *TLR ligands.* The PAMP recognized by TLR3 is double-stranded RNA (dsRNA). Against which pathogens would TLR3 be an effective immune receptor?

6. *Energetics and kinetics.* Suppose that the dissociation constant of an F_{ab}–hapten complex is 3×10^{-7} M at 25°C.

(a) What is the standard free energy of binding?

(b) Immunologists often speak of affinity (K_a), the reciprocal of the dissociation constant, in comparing antibodies. What is the affinity of this F_{ab}?

(c) The rate constant for the release of hapten from the complex is 120 s^{-1}. What is the rate constant for association? What does the magnitude of this value imply about the extent of structural change in the antibody on binding hapten?

7. *A brilliant emitter.* Certain naphthalene derivatives, such as the dansyl group, exhibit a weak yellow fluorescence when they are in a highly polar environment (such as water) and an intense blue fluorescence when they are in a markedly nonpolar environment (such as hexane). The binding of ε-dansyl-lysine to specific antibody is accompanied by a marked increase in its fluorescence intensity and a shift in color from yellow to blue. What does this finding reveal about the hapten–antibody complex?

8. *Miniantibody.* The F_{ab} fragment of an antibody molecule has essentially the same affinity for a monovalent hapten as does intact IgG.

(a) What is the smallest unit of an antibody that can retain the specificity and binding affinity of the whole protein?

(b) Design a compact single-chain protein that is likely to specifically bind antigen with high affinity.

9. *Turning on B cells.* B lymphocytes, the precursors of plasma cells, are triggered to proliferate by the binding of multivalent antigens to receptors on their surfaces. The cell-surface receptors are transmembrane immunoglobulins. Univalent antigens, in contrast, do not activate B cells.

(a) What do these findings reveal about the mechanism of B-cell activation?

(b) How might antibodies be used to activate B cells?

10. *An ingenious cloning strategy.* In the cloning of the gene for the α chain of the T-cell receptor, T-cell cDNAs were

hybridized with B-cell mRNAs. What was the purpose of this hybridization step? Can the principle be applied generally?

11. *Pathogen susceptibility.* Patients carrying specific mutations in the gene encoding the TLR4 protein are susceptible to infections from Gram-negative bacteria. Why are these patients vulnerable to this particular type of pathogen?

12. *Matchmaker, matchmaker.* Why is it important to match HLA alleles between donor and recipient in organ transplantation?

13. *Instruction.* Before the mechanism for generating antibody diversity had been established, a mechanism based on protein folding around an antigen was proposed, primarily by Linus Pauling. In this model, antibodies that had different specificities had the same amino acid sequence but were folded in different ways. Propose a test of this model.

14. *Dealing with nonsense.* Cells, including immune cells, degrade mRNA molecules in which no long open reading frame is present. The process is called nonsense-mediated RNA decay. Suggest a role for this process in immune cells.

15. *Down, but not out.* To understand the genes responsible for growth and infectivity in a disease-causing bacterial strain, you perform chemical mutagenesis on a culture of these bacteria. In the course of your investigation into the properties of the resulting mutants, you identify a set of mutant bacteria that is still viable but their virulence is significantly impaired. How might these mutants be useful for vaccine development?

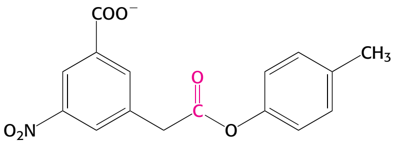
16. *Presentation.* The amino acid sequence of a small protein is

`MSRLASKNLIRSDHAGGLLQATYSAVSSIKNTMSFGAWSNAALNDSRDA`

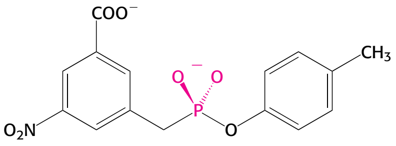
Predict the most likely peptide to be presented by the class I MHC molecule HLA-A2.

Mechanism Problem

17. *Catalytic antibody.* Antibody is generated against a transition state for the hydrolysis of the following ester.



Ester



Transition-state analog

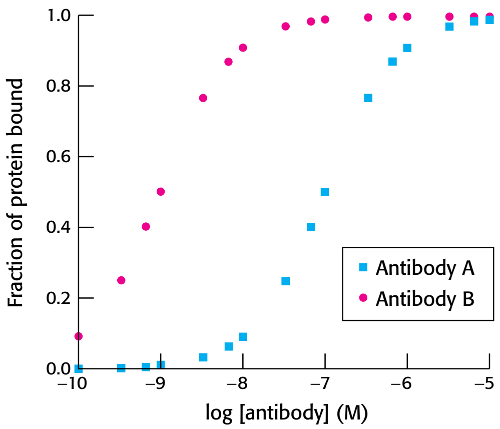
Some of these antibodies catalyze the hydrolysis of the ester. What amino acid residue might you expect to find in the binding site on the antibody?

Chapter Integration Problem

18. *Signaling.* Protein tyrosine phosphatases, such as the molecule CD45 expressed in both B cells and T cells, play important roles in activating such protein tyrosine kinases as Fyn and Lck, which are quite similar to Src. Suggest a mechanism for the activation of such protein kinases by the removal of a phosphoryl group from a phosphotyrosine residue.

Data Interpretation Problem

19. *Affinity maturation.* A mouse is immunized with an oligomeric human protein. Shortly after immunization, a cell line that expresses a single type of antibody molecule (antibody A) is derived. The ability of antibody A to bind the human protein is assayed, with the results shown in the graph below.



After repeated immunizations with the same protein, another cell line is derived that expresses a different antibody (antibody B). The results of analyzing the binding of antibody B to the protein also are shown. From these data, estimate

(a) the dissociation constant (K_d) for the complex between the protein and antibody A.

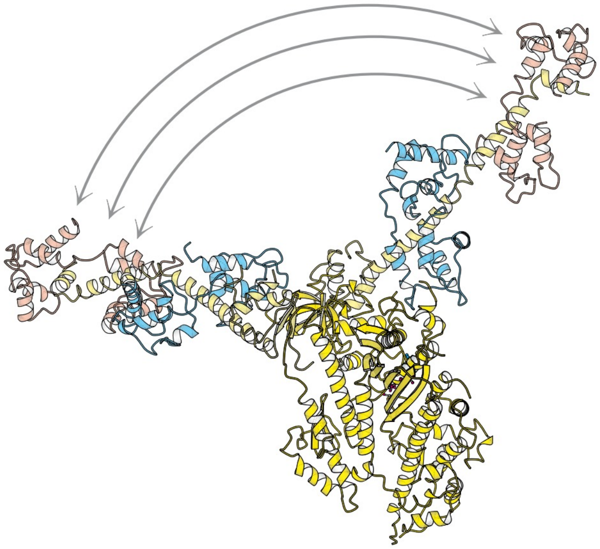
(b) the dissociation constant for the complex between the protein and antibody B.

Comparison of the amino acid sequences of antibody A and antibody B reveals them to be identical except for a single amino acid. What does this finding suggest about the mechanism by which the gene encoding antibody B was generated?

CHAPTER

35

Molecular Motors



The impressive muscles of the horse, like the muscles of all animals, are powered by the molecular-motor protein myosin. A part of myosin moves dramatically (as shown above) in response to ATP binding, hydrolysis, and product release, propelling myosin along an actin filament. This molecular movement is translated into movement of the entire animal, vividly depicted in da Vinci's rearing horse. [(Left) The Royal Collection © 2014 Her Majesty Queen Elizabeth II/Bridgeman Images.]

OUTLINE

35.1 Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily

35.2 Myosins Move Along Actin Filaments

35.3 Kinesin and Dynein Move Along Microtubules

35.4 A Rotary Motor Drives Bacterial Motion

Organisms, from human beings to bacteria, move to adapt to changes in their environments, navigating toward food and away from danger. Cells themselves are not static but are bustling assemblies of moving proteins, nucleic acids, and organelles. This motion is enabled by two elements: molecular-motor proteins and complex networks of filamentous proteins termed the *cytoskeleton* (Figure 35.1). The dynamic networks that determine the shape and mobility of calls are among the most active areas of investigation in modern cell biology. Remarkably, the fundamental biochemical mechanisms that produce contractions in our muscles are the same as those that propel organelles along the cytoskeleton. In fact, many of the proteins that play key roles in converting chemical energy into kinetic energy are members of the same protein family, the P-loop NTPases, the hugely important group of proteins that we first examined in Chapter 9. These molecular motors are homologous to proteins that we have encountered in other contexts, including the G proteins in protein synthesis, signaling, and other processes. Once again, we see the economy of evolution in adapting existing proteins to perform new functions.

Molecular motors operate by small increments, converting changes in protein conformation into directed motion. Orderly motion across distances requires a track that steers the motion of the motor assembly. Indeed, we have

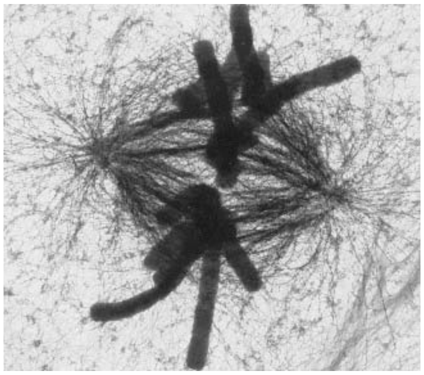


FIGURE 35.1 Motion within cells. This high-voltage electron micrograph shows the mitotic apparatus in a metaphase mammalian cell. The large cylindrical objects are chromosomes, and the threadlike structures stretched across the center are microtubules, key components of the skeleton. Microtubules serve as tracks for the molecular motors that move chromosomes. Many processes, including chromosome segregation in mitosis, depend on the action of molecular-motor proteins. [Courtesy of Dr. J. R. McIntosh.]

already encountered a class of molecular motors that utilize mechanisms that we will examine here—namely, the helicases that move along DNA during DNA replication (Section 28.1). The proteins on which we will focus in this chapter move along actin and microtubules—protein filaments composed of repeating subunits. The motor proteins cycle between forms having high or low affinity for the filament tracks in response to ATP binding and hydrolysis, enabling a bind, pull, and release mechanism that generates motion.

We will also consider a completely different strategy for generating motion, one used by bacteria such as *E. coli*. A set of flagella act as propellers, rotated by a motor in the bacterial cell membrane. This rotary motor is driven by a proton gradient across the membrane, rather than by ATP hydrolysis. The mechanism for coupling the proton gradient to rotatory motion is analogous to that used by the F_0 subunit of ATP synthase. Thus, both of the major modes for storing biochemical energy—namely, ATP and ion gradients—have been harnessed by evolution to drive organized molecular motion.

35.1 Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily

Eukaryotic cells contain three major families of motor proteins: myosins, kinesins, and dyneins. Members of each of these classes move along components of the cytoskeleton, but, at first glance, these protein families appear to be quite different from one another. *Myosin*, first characterized on the basis of its role in muscle, moves along filaments of the protein actin. Each molecule of muscle myosin consists of two copies each of a *heavy chain* with a molecular mass of 220 kDa, an *essential light chain*, and a *regulatory light chain*. The human genome encodes more than 40 distinct myosins; some function in muscle contraction, and others participate in a variety of other processes. *Kinesins,* which have roles in protein, mRNA, and vesicle transport as well as construction of the mitotic spindle and chromosome segregation, are generally dimers of two polypeptides. The human genome encodes more than 40 kinesins. *Dyneins* power the motion of cilia and flagella, and a general cytoplasmic dynein contributes to a variety of motions in all cells, including vesicle transport and various transport events in mitosis. Dyneins are enormous, with heavy chains of molecular mass greater than 500 kDa. The human genome encodes approximately 10 dyneins.

Initially, comparison of the amino acid sequences of myosins, kinesins, and dyneins did not reveal significant relationships between these protein families but, after their three-dimensional structures were determined, members of the myosin and kinesin families were found to have remarkable similarities. In particular, both myosin and kinesin contain P-loop NTPase cores homologous to those found in G proteins. Sequence analysis of the dynein heavy chain reveals it to be a member of the AAA subfamily of P-loop NTPases that we encountered in the context of the 19S proteasome (Section 23.2). Dynein has six sequences encoding such P-loop NTPase domains arrayed along its length, although only four actually appear to bind nucleotides. Thus, we expect similarities in the mechanisms of action, and we can draw on our knowledge of P-loop NTPases in general as we analyze the mechanisms of action of these motor proteins.

Molecular motors are generally oligomeric proteins with an ATPase core and an extended structure

Let us first consider the structure of myosin, which we examined briefly in Chapter 9. The results of electron microscopic studies of skeletal-muscle

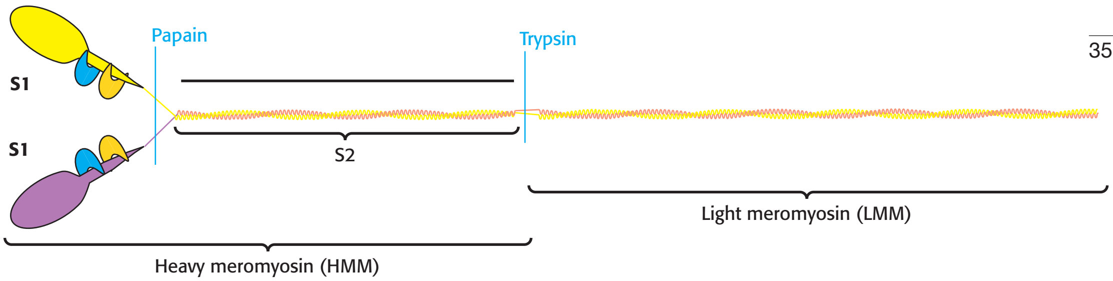

Papain
Trypsin
S1
S1
S2
Light meromyosin (LMM)
Heavy meromyosin (HMM)

FIGURE 35.2 Myosin dissection. Treatment of muscle myosin with proteases forms stable fragments, including subfragments S1 and S2 and light meromyosin. Each S1 fragment includes a head (shown in yellow or purple) from the heavy chain and one copy of each light chain (shown in blue and orange).

myosin show it to be a two-headed structure linked to a long stalk (Figure 9.42). The treatment of myosin with trypsin and papain results in the formation of four fragments: two S1 fragments; heavy meromyosin (HMM) which consists of the S1 fragments and an additional region termed S2; and a fragment called light meromyosin (LMM; Figure 35.2). Each *S1 fragment* corresponds to one of the heads of the intact structure and includes 850 amino-terminal amino acids from one of the two heavy chains as well as one copy of each of the light chains. Examination of the structure of an S1 fragment at high resolution reveals the presence of the P-loop NTPase-domain core that is the site of ATP binding and hydrolysis (Figure 35.3). We examined the structure and mechanism of action of this motor domain in Chapter 9.

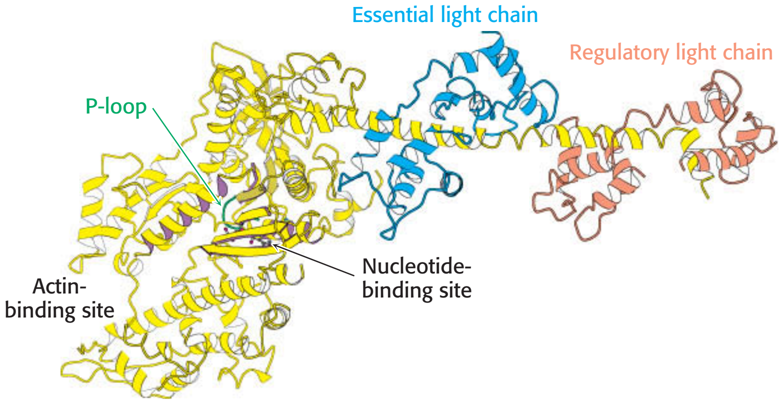

Essential light chain
Regulatory light chain
P-loop
Actin-
binding site
Nucleotide-
binding site

FIGURE 35.3 Myosin structure at high resolution. The structure of the S1 fragment from muscle myosin reveals the presence of a P-loop NTPase domain (shaded in purple). *Notice* that an α helix extending from this domain is the binding site for the two light chains. [Drawn from 1DFL.pdb.]

Extending away from this structure is a long α helix from the heavy chain. This helix is the binding site for the two light chains. The light chains are members of the EF-hand family, similar to calmodulin (Figure 14.17), although most of the EF hands in light chains do not bind metal ions (Figure 35.4). Like calmodulin, these proteins wrap around an α helix, serving to thicken and

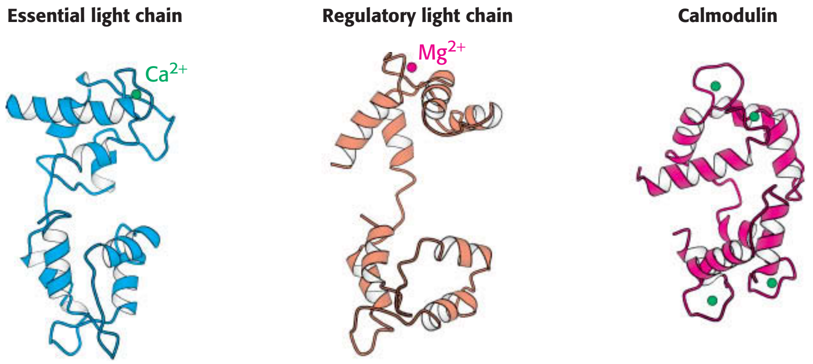

Essential light chain
Ca2+
Regulatory light chain
Mg2+
Calmodulin

FIGURE 35.4 Myosin light chains. The structures of the essential and regulatory light chains of muscle myosin are compared with the structure of calmodulin. *Notice* the similarities in the structures that allow each of these homologous proteins to bind an α helix (not shown) by wrapping around it. [Drawn from 1DFL.pdb and 1CM1.pdb.]

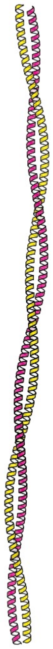


FIGURE 35.5 **Myosin two-stranded coiled coil.** The two α helices form left-handed supercoiled structures that spiral around each other. Such structures are stabilized by hydrophobic residues at the contact points between the two helices. [Drawn from 2TMA.pdb.]

stiffen it. The remaining fragments of myosin—S2 and light meromyosin—are largely α helical, forming two-stranded coiled coils created by the remaining lengths of the two heavy chains wrapping around each other (Figure 35.5). These structures, together extending approximately 1700 Å, link the myosin heads to other structures. In muscle myosin, several LMM domains come together to form higher-order bundles.

Conventional kinesin (kinesin 1), the first kinesin discovered, has several structural features in common with myosin. The dimeric protein has two heads connected by an extended structure (Figure 35.6). The size of the head domain is approximately one-third that of myosin. Determination of the three-dimensional structure of a kinesin fragment revealed that the head domain also is built around a P-loop NTPase core (Figure 35.7). The myosin domain is so much larger than that of kinesin because of two large insertions in the myosin domain that bind to actin filaments. For conventional kinesin, a region of approximately 500 amino acids extends from the head domain. Like the corresponding region in myosin, the extended part of kinesin forms an α-helical coiled coil. Conventional kinesin also has light chains, but, unlike those of myosin, these light chains bind near the carboxyl terminus of the heavy chain and are thought to link the motor to intracellular cargo.

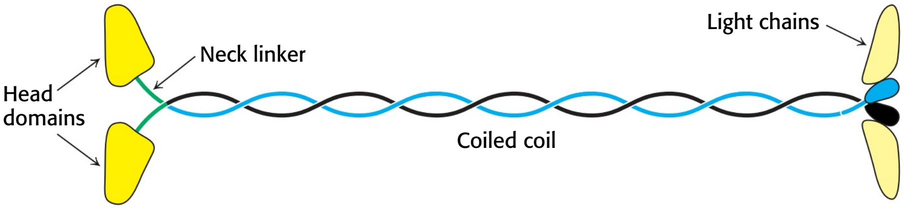


FIGURE 35.6 **Structure of kinesin.** The elongated structure has the head domains at one end and the cargo-binding domains at the other, linked by a long coiled-coil region.

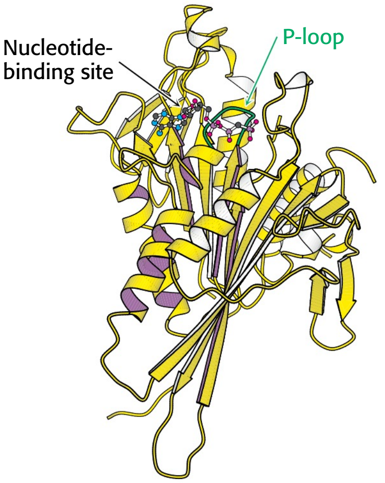


FIGURE 35.7 **Structure of head domain of kinesin at high resolution.** *Notice* that the head domain of kinesin has the structure of a P-loop NTPase core (indicated by purple shading). [Drawn from 1I6I.pdb.]

Dynein has a significantly different structure than myosin or kinesin. As noted earlier, the dynein heavy chain includes six regions that are homologous to the AAA subfamily of ATPase domains. A model for the structure of the motor domain of dynein was generated based on the structures of other AAA ATPases. This model has been confirmed and extended by more recent structure determinations of dynein itself (Figure 35.8). The head domain is appended to a region of approximately 1300 amino acids that forms an extended structure that links dynein units together to form oligomers and interacts with other proteins.

Although the structures of these three classes of molecular motors have significant differences, some common features emerge. Almost all structures are dimeric with two head domains, have regions of extended but quite rigid structures, and have regions for interacting with other proteins. As we shall see, these structures are suitable for actions that resemble climbing a rope, hand over hand. The regions for interacting with other proteins represent the grasping hands, the extended structures represent the arms that act as levers to promote larger-scale motion, and the head domains are the engines that provide the necessary mechanical energy.

ATP binding and hydrolysis induce changes in the conformation and binding affinity of motor proteins

In Chapter 9, we examined the conformational changes that take place in the myosin ATPase domain from the slime mold *Dictyostelium*. The structures

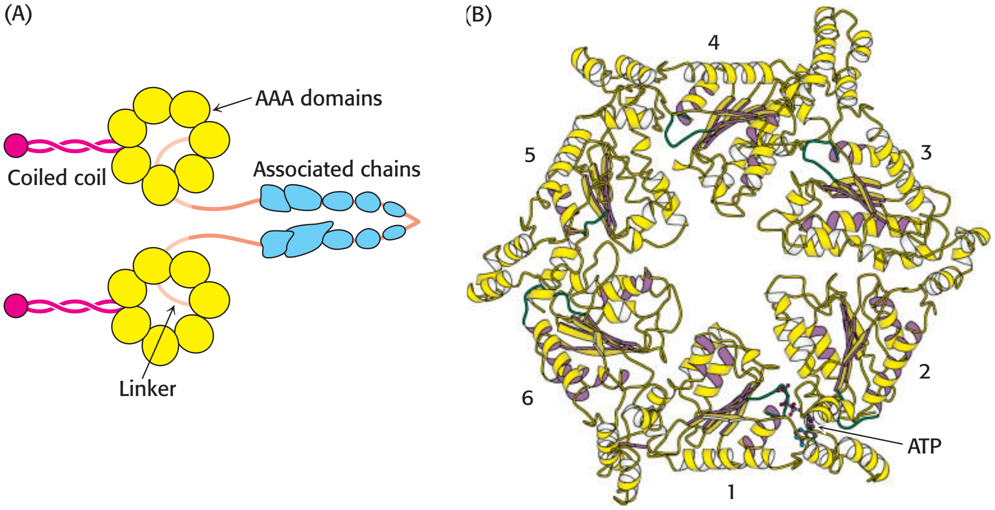


FIGURE 35.8 Dynein structure. (A) The overall structure of dynein. (B) A model of the motor domain of dynein. *Notice* the six P-loop NTPase domains, some of which bind to and hydrolyze ATP. [Drawn from 1HN5.pdb.]

of myosin ATP domains from other sources have been elucidated in a variety of forms as well. The S1 fragment of myosin from scallop muscle provides a striking example of the changes observed (Figure 35.9). The structure of this S1 fragment has been determined in a number of forms including that without bound nucleotide and that bound to a complex formed of ADP and vanadate (VO_4^{3-}), which, as mentioned in Chapter 9, is an analog of the ATP-hydrolysis transition state. The long helix that binds the light chains (hereafter referred to as the *lever arm*) protrudes outward from the head domain. Comparison of the structures reveals that the lever arm has rotated by nearly 90 degrees in the ADP–VO_4^{3-} complex compared with its position in the nucleotide-free form. How does the species in the nucleotide-binding site cause this dramatic transition? Two regions around the nucleotide-binding site (termed *switch I* and *switch II*) tightly conform to the shape of the γ-phosphoryl group of the ATP analog and adopt a looser

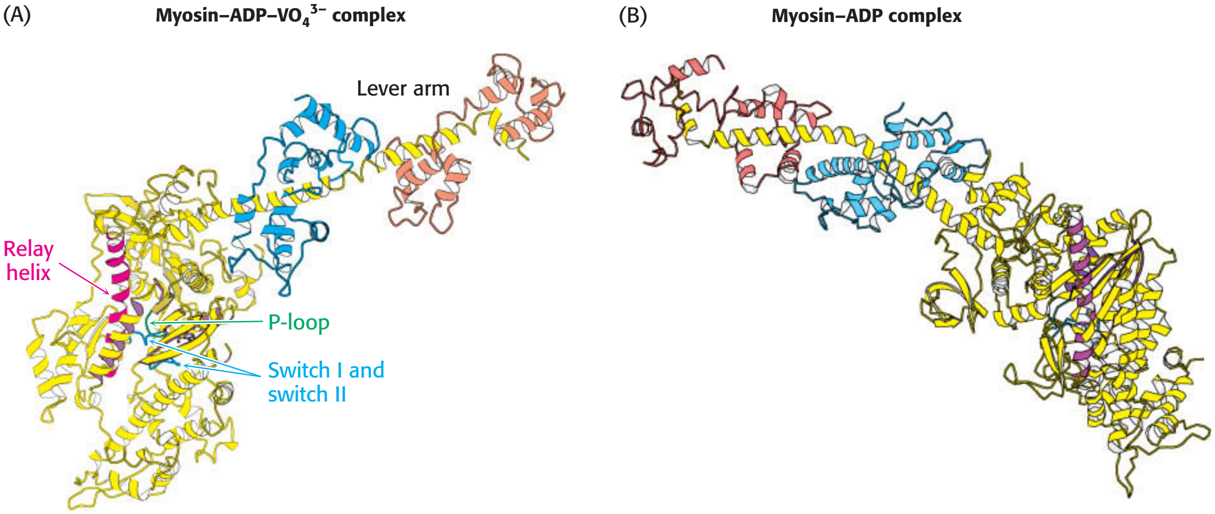


FIGURE 35.9 Lever-arm motion. Two forms of the S1 fragment of scallop-muscle myosin. *Notice* the dramatic conformational changes when the identity of the bound nucleotide changes from the ADP–VO_4^{3-} complex (A) to the nucleotide-free form (B) or vice versa, including a nearly 90-degree reorientation of the lever arm. [Drawn from 1DFL.pdb and 1SR6.pdb.]

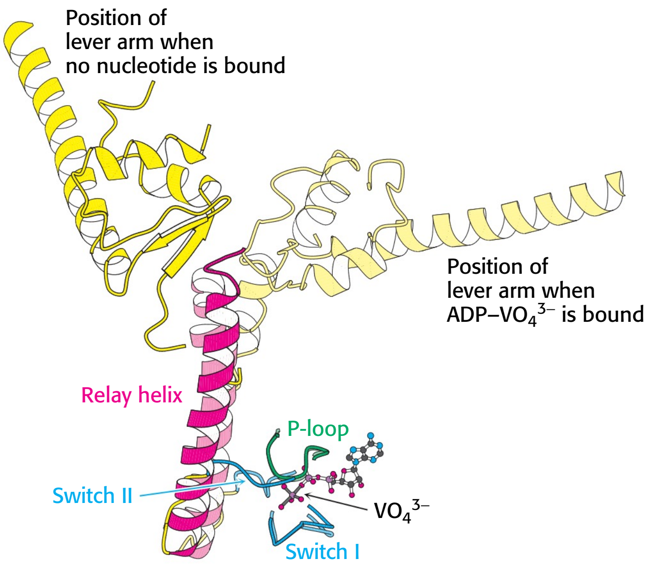

Position of lever arm when no nucleotide is bound
Position of lever arm when ADP–VO_4^{3-} is bound
Relay helix
P-loop
Switch II
VO_4^{3-}
Switch I

FIGURE 35.10 Relay helix. A superposition of key elements in two forms of scallop myosin reveals the structural changes that are transmitted by the relay helix from the switch I and switch II loops to the base of the lever arm. *Notice* that the switch I and switch II loops interact with VO_4^{3-} in the position that would be occupied by the γ-phosphoryl group of ATP. The structure of the myosin–ADP–VO_4^{3-} complex is shown in lighter colors. [Drawn from 1DFL.pdb and 1SR6.pdb.]

conformation when the γ-phosphoryl group is absent (Figure 35.10). This conformational change allows a long α helix (termed the *relay helix*) to adjust its position. The carboxyl-terminal end of the relay helix interacts with structures at the base of the lever arm, and so a change in the position of the relay helix leads to a reorientation of the lever arm.

Analogous conformational changes take place in kinesin. The kinesins also have a relay helix that can adopt different configurations when kinesin binds different nucleotides. Kinesin lacks an α-helical lever arm, however. Instead, a relatively short segment termed the *neck linker* changes conformation in response to nucleotide binding (Figure 35.11). The neck linker binds to the head domain of kinesin when ATP is bound but is released when the nucleotide-binding site is vacant or occupied by ADP.

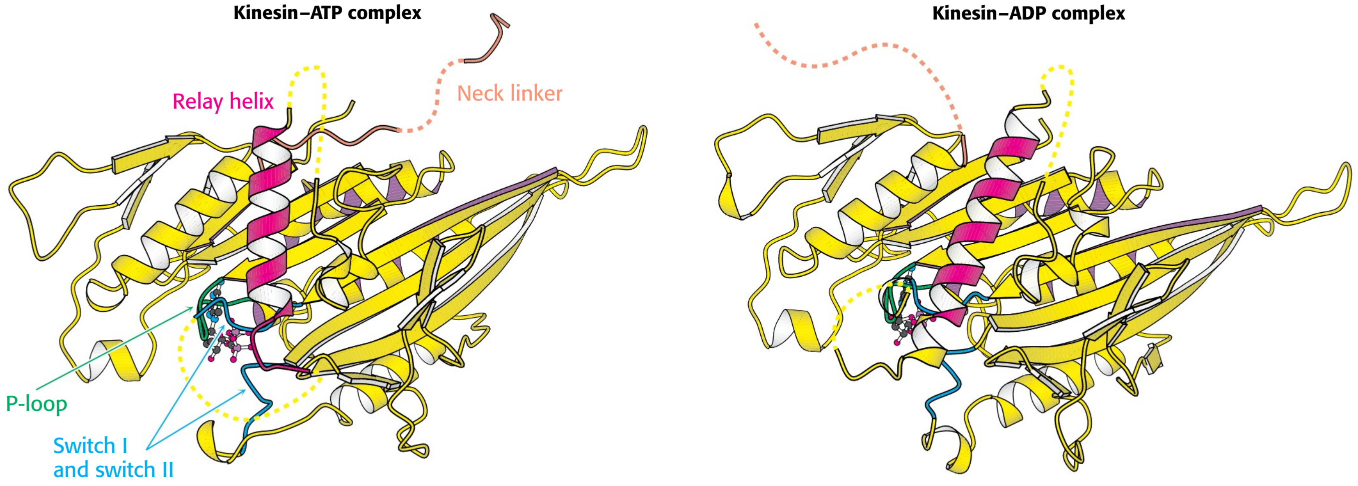
35.2 Myosins Move Along Actin Filaments

Myosins, kinesins, and dyneins move by cycling between states with different affinities for the long, polymeric macromolecules that serve as tracks along which they move. For myosin, the molecular track is a polymeric form of *actin,* a 42-kDa protein that is among the most abundant in cells, typically accounting for as much as 10% of the protein in eukaryotic cells. We begin with a general discussion of the polymeric structure of actin and its assembly. We then examine the interactions between myosin and actin, including both structure and the dynamic interactions between these two proteins. Finally, we turn to the structure of muscle and the roles of myosin and actin in muscle contraction.

Actin is a polar, self-assembling, dynamic polymer

The structure of the actin monomer has been determined to atomic resolution by x-ray crystallography and has been used to interpret the structure of actin filaments, already somewhat understood through electron microscopy studies at lower resolution. Each actin monomer comprises four domains

FIGURE 35.11 Neck linker. A comparison of the structures of a kinesin bound to ADP and bound to an ATP analog. *Notice* that the neck linker (orange), which connects the head domain to the remainder of the kinesin molecule, is bound to the head domain in the presence of the ATP analog but is free in the presence of ADP only. [Drawn from 1I6I.pdb and 1I5S.pdb.]



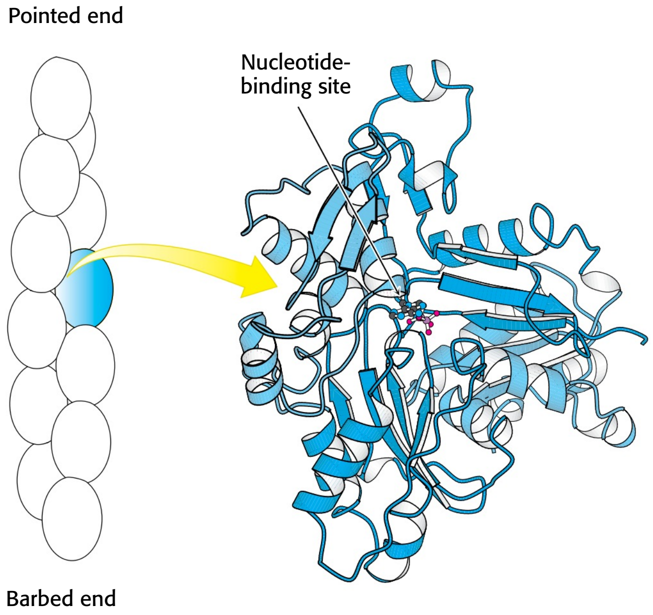


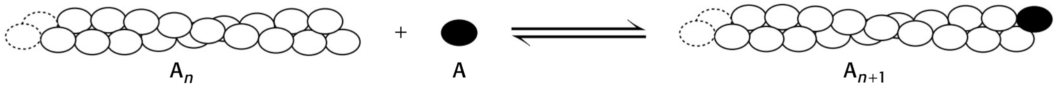
FIGURE 35.12 Actin structure. (Left) Schematic view of actin monomers (one in blue) of an actin filament. (Right) The domains in the four-domain structure of an actin monomer are identified by different shades of blue. *Notice* the nucleotide-binding site at the center of the structure. [Drawn from 1J6Z.pdb.]

(Figure 35.12). These domains come together to surround a bound nucleotide, either ATP or ADP. The ATP form can be converted into the ADP form by hydrolysis.

Actin monomers (often called *G-actin* for globular) come together to form actin filaments (often called *F-actin;* see Figure 35.12). F-actin has a helical structure; each monomer is related to the preceding one by a translation of 27.5 Å and a rotation of 166 degrees around the helical axis. Because the rotation is nearly 180 degrees, F-actin resembles a two-stranded cable. Note that each actin monomer is oriented in the same direction along the F-actin filament, and so the structure is polar, with discernibly different ends. One end is called the barbed (plus) end, and the other is called the pointed (minus) end. The names "barbed" and "pointed" refer to the appearance of an actin filament when myosin S1 fragments are bound to it.

How are actin filaments formed? Like many biological structures, actin filaments self-assemble; that is, under appropriate conditions, actin monomers will come together to form well-structured, polar filaments. The aggregation of the first two or three monomers to form a filament is highly unfavorable. Thus, specialized protein complexes, including one called Arp2/3, serve as nuclei for actin assembly in cells. Once such a filament nucleus exists, the addition of subunits is more favorable. Let us consider the polymerization reaction in more detail. We designate an actin filament with n subunits A_n. This filament can bind an additional actin monomer, A, to form A_{n+1}.

$$K_d = \frac{[A_n][A]}{[A_{n+1}]}$$



The dissociation constant, K_d, for this reaction, defines the monomer concentrations at which the polymerization reaction will take place, because the concentration of polymers of length $n + 1$ will be essentially equal to that for polymers of length n. Thus,

$$[A_n] \sim [A_{n+1}] \qquad \text{and} \qquad K_d = \frac{[A_n][A]}{[A_{n+1}]} \sim [A]$$

In other words, the polymerization reaction will proceed until the monomer concentration is reduced to the value of K_d. If the monomer concentration is below the value of K_d, the polymerization reaction will not proceed at all; indeed, existing filaments will depolymerize until the monomer concentration reaches the value of K_d. Because of these phenomena, K_d is referred to as the *critical concentration* for the polymer. Recall that actin contains a nucleotide-binding site that can contain either ATP or ADP. The critical concentration for the actin–ATP complex is approximately 20-fold lower than that for the actin–ADP complex; actin–ATP polymerizes more readily than does actin–ADP.

Actin filaments inside cells are highly dynamic structures that are continually gaining and losing monomers. Nucleation by complexes such as Arp2/3 can initiate the polymerization of actin–ATP. In contrast, the hydrolysis of bound ATP to ADP favors actin *de*polymerization. This reaction acts as a timer to make actin filaments kinetically unstable as ATP hydrolyzes over time, making the filaments more unstable. Proteins that bind actin monomers or promote the severing of actin filaments also play roles. Polymerization reactions can exert force, pushing or pulling on cell membranes. *Regulated actin polymerization is central to the changes in cell shape associated with cell motility in amebae as well as in human cells such as macrophages.*

Myosin head domains bind to actin filaments

It has not been possible to determine the in vivo structure of a complex between actin and myosin at sufficiently high resolution to discern molecular details. However, treatment of actin filaments with myosin S1 fragments in the absence of ATP results in a complex referred to as *decorated actin* for which the structure has been determined by cryoelectron microscopy to a resolution of 13 Å. Although a structure at this resolution alone would not be adequate to observe molecular details, superimposition of the high-resolution structures of actin monomers and the myosin S1 fragment on the structure of decorated actin can be a source of insight into the details of its structure (Figure 35.13). The myosin head domain is in a conformation close to that observed for the nucleotide-free form. This structure also reveals the interaction surfaces between myosin and actin. The modeling suggests that the myosin head-domain conformation changes somewhat to increase its interaction with the actin filament. These conformational changes result in a slight opening of the nucleotide-binding site in myosin. This observation has implications for the mechanism by which myosin moves along actin filaments.

Motions of single motor proteins can be directly observed

Now that we understand the conformational changes behind myosin's action, we can explore how myosin "walks" along its actin track. Studies of *single myosin molecules* moving relative to actin filaments have been sources of deep insight into the mechanisms underlying muscle contraction and

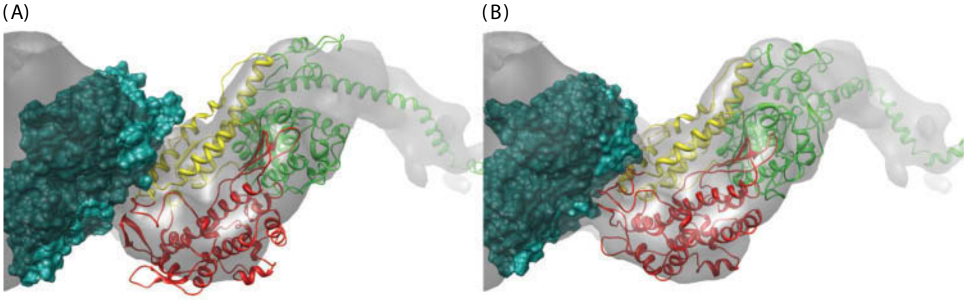


FIGURE 35.13 The structure of myosin bound to actin. (A) The gray surface represents the structure observed by cryoelectron microscopy, with the green space-filling model representing one actin subunit. The ribbon diagram shows the structure of the S1 fragment of myosin docked into the cryoelectron microscopic structure. *Notice* that some of the myosin structure lies outside the gray surface. (B) The structure after the myosin S1 fragment has been allowed to adjust to more closely match the structure observed by cryoelectron microscopy. *Notice* that the myosin structure now more closely matches the gray surface. [M. Lorenz and K. Holmes. *PNAS* 2010 (107) 12529–12534. Copyright 2010 National Academy of Sciences.]

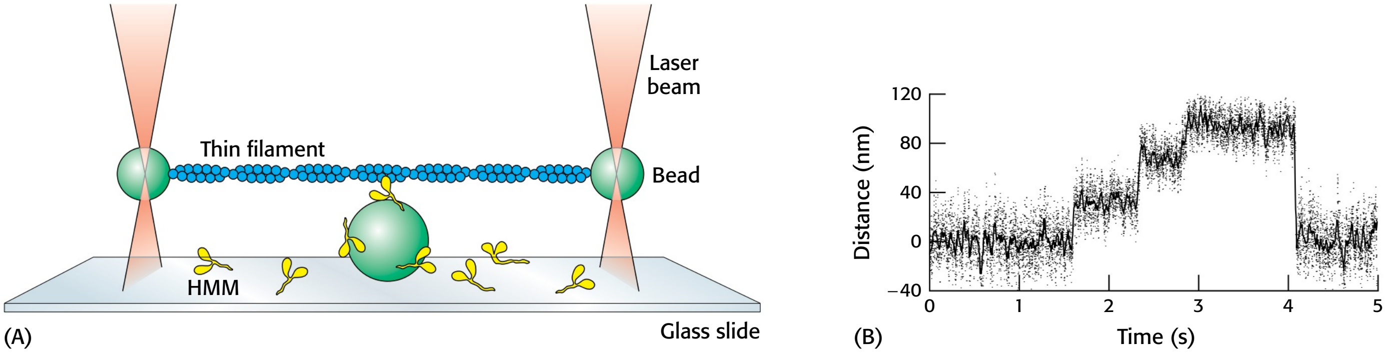


FIGURE 35.14 Watching a single motor protein in action. (A) An actin filament (blue) is placed above a heavy meromyosin (HMM) fragment (yellow) that projects from a bead on a glass slide. A bead attached to each end of the actin filament is held in an optical trap produced by a focused, intense infrared laser beam (orange). The position of these beads can be measured with nanometer precision. (B) Recording of the displacement of an actin filament due to a myosin derivative attached to a bead, influenced by the addition of ATP. Note the fairly uniform step sizes that are observed. [(A) Information from J. T. Finer, R. M. Simmons, and J. A. Spudich, *Nature* 368:113–119, 1994. (B) Data from R. S. Rock, M. Rief, A. D. Metra, and J. A. Spudich, *Methods* 22:378–381, 2000.]

other complex processes. A powerful tool for these studies, called an *optical trap,* relies on highly focused laser beams (Figure 35.14). Small beads can be caught in these traps and held in place in solution.

The position of the beads can be monitored with nanometer precision. James Spudich and his coworkers designed an experimental arrangement consisting of an actin filament that had a bead attached to each end. Each bead could be caught in an optical trap (one at each end of the filament) and the actin filament could be pulled taut over a microscope slide containing other beads that had been coated with fragments of myosin such as the heavy meromyosin fragment (Figure 35.14). On the addition of ATP, transient displacements of the actin filament were observed along its long axis. The size of the displacement steps was fairly uniform with an average size of 11 nm (110 Å).

The results of these studies, performed in the presence of varying concentrations of ATP, are interpreted as showing that individual myosin heads bind the actin filament and undergo a conformational change (the *power stroke*) that pulls the actin filament, leading to the displacement of the beads. After a period of time, the myosin head releases the actin, which then snaps back into place.

Phosphate release triggers the myosin power stroke

How does ATP hydrolysis drive the power stroke? A key observation is that the addition of ATP to a complex of myosin and actin results in the dissociation of the complex. Thus, ATP binding and hydrolysis cannot be directly responsible for the power stroke. We can combine this fact with the structural observations described earlier to construct a mechanism for the motion of myosin along actin (Figure 35.15). Let us begin with nucleotide-free myosin bound to actin. The binding of ATP to actin results in the dissociation of myosin from actin. With ATP bound and free of actin, the myosin domain can undergo the conformational change associated with the formation of the transition state for ATP hydrolysis. This conformational change results in the reorientation of the lever arm. In this form, the myosin head can dock onto the actin filament; phosphate is released with concomitant motion of the lever arm. *This conformational change represents the power stroke and moves the body of the myosin molecule relative to the actin filament by approximately 110 Å*. The release of ADP completes the cycle.

Muscle is a complex of myosin and actin

The mechanism of moving a single myosin molecule relative to an actin filament explains how muscles contract. Vertebrate muscle that is under voluntary control, such as the biceps and triceps in your upper arm, has a banded (striated) appearance when examined under a light microscope. It consists of multinucleated cells that are bounded by an electrically excitable plasma

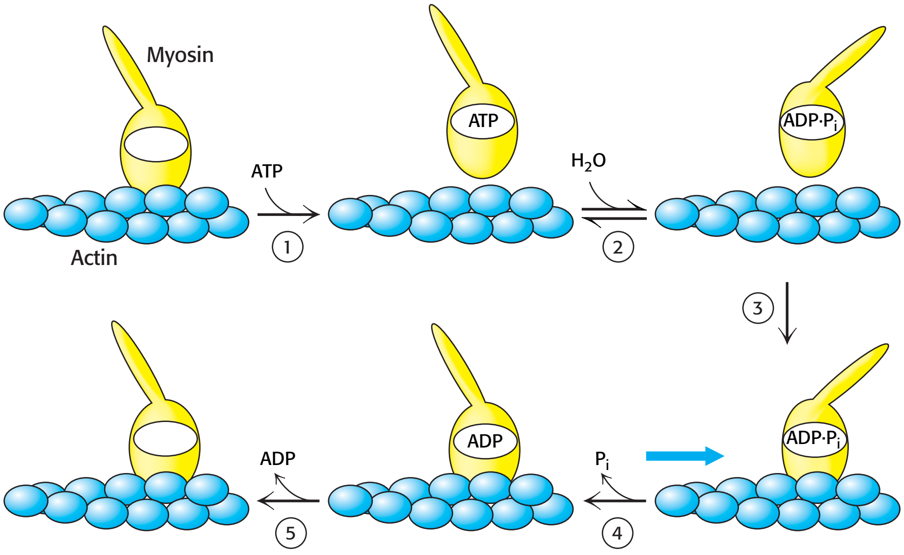


FIGURE 35.15 Myosin motion along actin. A myosin head (yellow) in the apo form is bound to an actin filament (blue). The binding of ATP (1) results in the release of myosin from actin. The reversible hydrolysis of ATP bound to myosin (2) can result in the reorientation of the lever arm. With ATP hydrolyzed but still bound to actin, myosin can bind actin (3). The release of P_i (4) results in the reorientation of the lever arm and the concomitant motion of actin relative to myosin. The release of ADP (5) completes the cycle.

membrane. A muscle cell contains many parallel *myofibrils,* each about 1 μm in diameter. The functional unit, called a *sarcomere,* typically repeats every 2.3 μm (23,000 Å) along the fibril axis in relaxed muscle (Figure 35.16). A dark *A band* and a light *I band* alternate regularly. The central region of the A band, termed the *H zone,* is less dense than the rest of the band. The I band is bisected by a very dense, narrow *Z line*.

The underlying molecular plan of a sarcomere is revealed by cross sections of a myofibril. These cross sections show the presence of two kinds of interacting protein filaments. The *thick filaments* have diameters of about 15 nm (150 Å) and consist primarily of myosin. The *thin filaments* have diameters of approximately 8 nm (80 Å) and consist of actin as well as *tropomyosin* and the *troponin complex*. Muscle contraction is achieved through the sliding of the thin filaments along the length of the thick filaments, driven by the hydrolysis of ATP (Figure 35.17).

To form the thick filaments, myosin molecules self-assemble into thick bipolar structures, with the myosin heads protruding at both ends of a bare region in the center (Figure 35.18A). Approximately 500 head domains line the surface of each thick filament. Each head-rich region associates with two actin filaments, one on each side of the myosin molecules (Figure 35.18B). The interaction of individual myosin heads with actin units creates the sliding force that gives rise to muscle contraction.

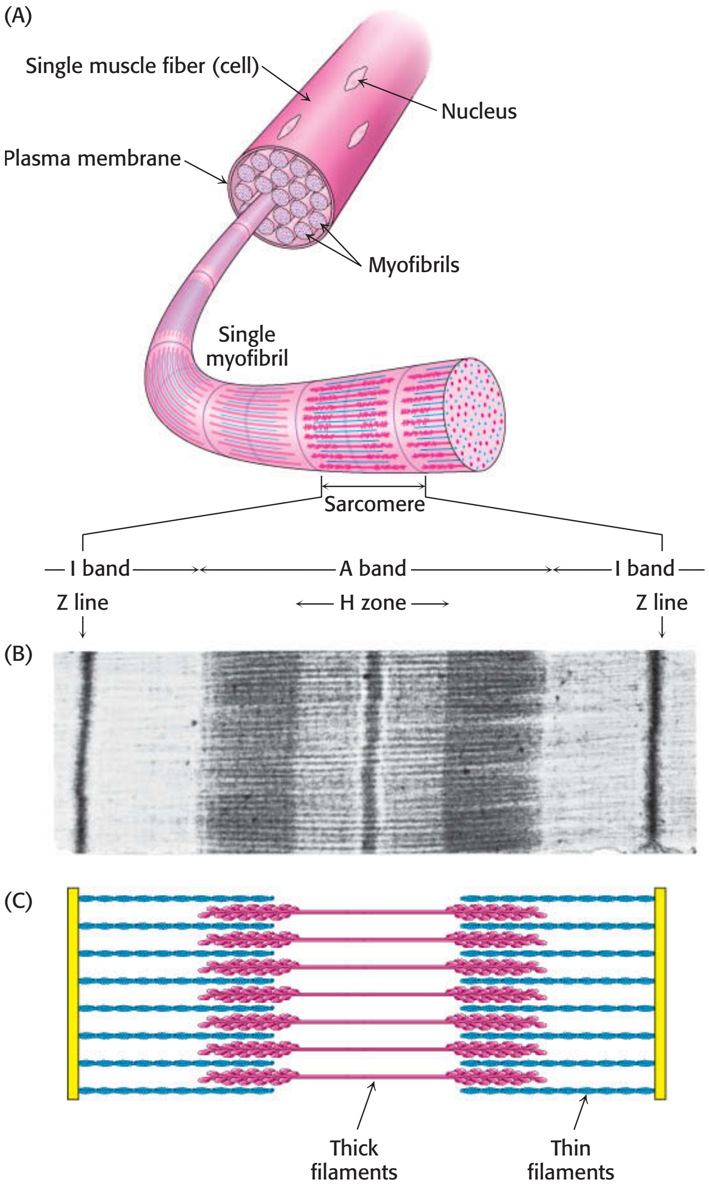


FIGURE 35.16 Sarcomere. (A) Structure of muscle cell and myofibril containing sarcomeres. (B) Electron micrograph of a longitudinal section of a skeletal-muscle myofibril showing a single sarcomere. (C) Schematic representation of the sarcomere corresponding to the regions in the micrograph. [(B) Courtesy of Dr. Hugh Huxley.]

Tropomyosin and the troponin complex regulate this sliding in response to nerve impulses. Under resting conditions, tropomyosin blocks the intimate interaction between myosin and actin. A nerve impulse leads to an increase in calcium ion concentration within the muscle cell. A component of the troponin complex senses the increase in Ca^{2+} and, in response, relieves the inhibition of myosin–actin interactions by tropomyosin.

How does the myosin reaction cycle apply to muscle contraction? Recall that hundreds of head domains project from the ends of each thick filament. The head domains are paired in myosin dimers, but the two heads within each dimer act independently. Actin filaments associate with each head-rich region, with the barbed ends of actin toward the Z line. In the presence of normal levels of ATP, most of the myosin heads are detached from actin. Each head can independently hydrolyze ATP, bind to actin, release P_i, and undergo its power stroke. Because few other heads are attached, the actin filament is relatively free to slide. Each head cycles approximately five times per second with a movement of 110 Å per cycle. However, when hundreds of heads are interacting with the same actin filament, the overall rate of movement of myosin relative to the actin filament may reach 80,000 Å per second, allowing a sarcomere to contract from its fully relaxed to its fully contracted form rapidly. Having many myosin heads briefly and independently attaching and moving an actin filament allows for much greater speed than could be achieved by a single motor protein.

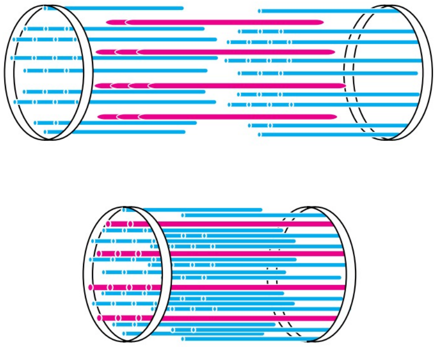


FIGURE 35.17 Sliding-filament model. Muscle contraction depends on the motion of thin filaments (blue) relative to thick filaments (red). [Information from H. E. Huxley. The mechanism of muscular contraction. Copyright © 1965 by Scientific American, Inc. All rights reserved.]

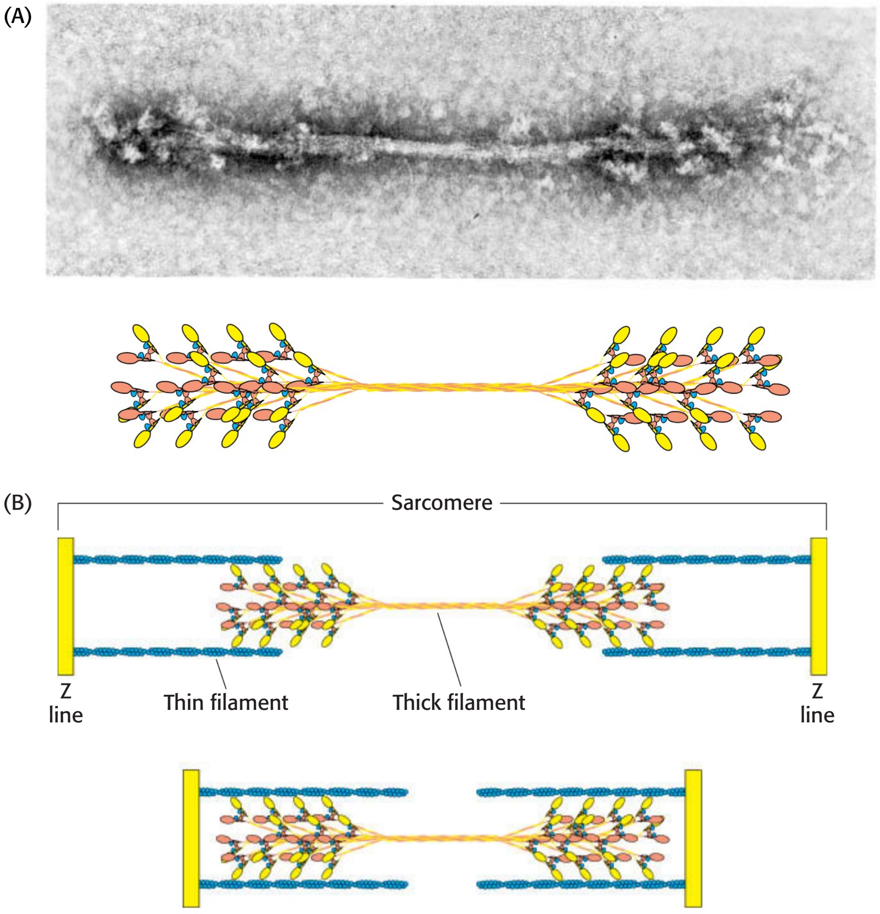


FIGURE 35.18 Thick filament. (A) An electron micrograph of a reconstituted thick filament reveals the presence of myosin head domains at each end and a relatively narrow central region. A schematic view below shows how myosin molecules come together to form the thick filament. (B) A diagram showing the interaction of thick and thin filaments in skeletal-muscle contraction. [(A, top) Courtesy of Dr. Hugh Huxley.]

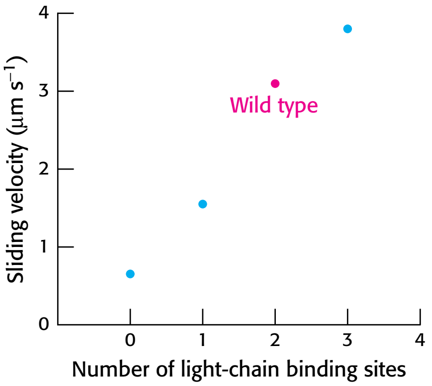


FIGURE 35.19 Myosin lever-arm length. Examination of the rates of actin movement supported by a set of myosin mutants with different numbers of light-chain binding sites revealed a linear relation; the greater the number of light-chain binding sites (and, hence, the longer the lever arm), the faster the sliding velocity. [Data from T. Q. P. Uyeda, P. D. Abramson, and J. A. Spudich, *Proc. Natl. Acad. Sci. U. S. A.* 93:4459–4464, 1996.]

The length of the lever arm determines motor velocity

A key feature of myosin motors is the role of the lever arm as an amplifier. The lever arm amplifies small structural changes at the nucleotide-binding site to achieve the 110-Å movement along the actin filament that takes place in each ATP hydrolysis cycle. A strong prediction of the mechanism proposed for the movement of myosin along actin is that the length traveled per cycle should depend on the length of this lever arm. Thus, the length of the lever arm should influence the overall rate at which actin moves relative to a collection of myosin heads.

This prediction was tested with the use of mutated forms of myosin with lever arms of different lengths. The lever arm in muscle myosin includes binding sites for two light chains (Section 35.1). Thus, investigators shortened the lever arm by deleting the sequences that correspond to one or both of these binding sites. They then examined the rates at which actin filaments were transported along collections of these mutated myosins (Figure 35.19). As predicted, the rate decreased as the lever arm was shortened. A mutated form of myosin with an unusually long lever arm was generated by inserting 23 amino acids corresponding to the binding site for an additional regulatory light chain. Remarkably, this form was found to support actin movement that was *faster than that of the wild-type protein*. These results strongly support the proposed role of the lever arm in contributing to myosin motor activity.

35.3 Kinesin and Dynein Move Along Microtubules

In addition to actin, the cytoskeleton includes other components, notably intermediate filaments and microtubules. Microtubules serve as tracks for two classes of motor proteins—namely, kinesins and dyneins. Kinesins moving along microtubules usually carry cargo such as organelles and vesicles from the center of a cell to its periphery. Dyneins are important in sliding microtubules relative to one another during the beating of cilia and flagella on the surfaces of some eukaryotic cells. Additionally, dynein carries cargo from the cell periphery to the cell center.

Some members of the kinesin family are crucial to the transport of organelles and other cargo to nerve endings at the peripheries of neurons. It is not surprising, then, that mutations in these kinesins can lead to nervous system disorders. For example, mutations in a kinesin called KIF1β can lead to the most common peripheral neuropathy (weakness and pain in the hands and feet), Charcot-Marie-Tooth disease, which affects 1 in 2500 people. A glutamine-to-leucine mutation in the P-loop of the motor domain of this kinesin has been found in some affected persons. Knockout mice with a disruption of the orthologous gene have been generated. Mice heterozygous for the disruption show symptoms similar to those observed in human beings; homozygotes die shortly after birth. Mutations in other kinesin genes have been linked to human spastic paraplegia. In these disorders, defects in kinesin-linked transport may impair nerve function directly, and the decrease in the activity of specific neurons may lead to other degenerative processes.

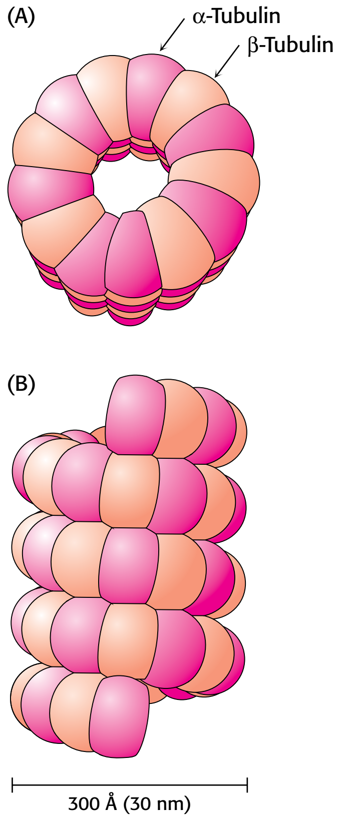


FIGURE 35.20 Microtubule structure. Schematic views of the helical structure of a microtubule. α-Tubulin is shown in dark red and β-tubulin in light red. (A) Top view. (B) Side view.

Microtubules are hollow cylindrical polymers

Microtubules are a major component of the cytoskeleton. They are built from two kinds of homologous 50-kDa subunits, α- and β-tubulin, which assemble in a helical array of alternating tubulin types to form the wall of a hollow cylinder (Figure 35.20). Alternatively, a microtubule can be regarded as 13 protofilaments that run parallel to its long axis. The outer diameter of a microtubule is 30 nm, much larger than that of actin (8 nm). Like actin, microtubules are polar structures. The minus end of a microtubule is anchored near the center of a cell, whereas the plus end extends toward the cell surface.

Microtubules are also key components of cilia and flagella present on some eukaryotic cells. For example, sperm propel themselves through the motion of flagella containing microtubules. The microtubules present in these structures adopt a common architecture (Figure 35.21). A bundle of microtubules called an *axoneme* is surrounded by a membrane contiguous with the plasma membrane. The axoneme is composed of a peripheral group of nine microtubule pairs surrounding two singlet microtubules. This recurring motif is often called a *9 + 2 array.* Dynein drives the motion of one member of each outer pair relative to the other, causing the overall structure to bend.

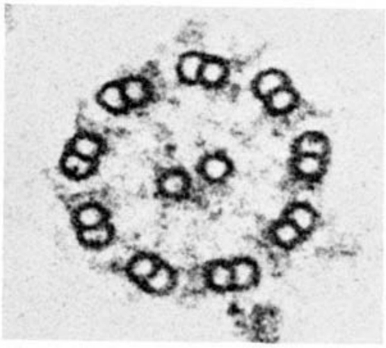


FIGURE 35.21 Microtubule arrangement. Electron micrograph of a cross section of a flagellar axoneme shows nine microtubule doublets surrounding two singlets. [Courtesy of Dr. Joel Rosenbaum.]

Microtubules are important in determining the shapes of cells and in separating daughter chromosomes in mitosis. They are highly dynamic structures that grow through the addition of α- and β-tubulin to the ends of existing structures. Like actin, *tubulins* bind and hydrolyze nucleoside triphosphates, although, for tubulin, the nucleotide is GTP rather than ATP. The critical concentration for the polymerization of the GTP forms of tubulin is lower than that for the GDP forms. Thus, a newly formed microtubule consists primarily of GTP-tubulins. Through time, the GTP is hydrolyzed to GDP. The GDP-tubulin subunits in the interior length of a microtubule remain stably polymerized, whereas GDP subunits exposed at an end have a strong tendency to dissociate. Marc Kirschner and Tim Mitchison found that some microtubules in a population lengthen while others simultaneously shorten. This property, called *dynamic instability,* arises from random fluctuations in the number of GTP- or GDP-tubulin subunits at the plus end of the polymer. The dynamic character of microtubules is crucial for processes such as mitosis, which require the assembly and disassembly of elaborate microtubule-based structures such as spindle fibers.

The structure of a tubulin dimer was determined at high resolution by electron crystallographic methods (Figure 35.22). As expected from their 40% sequence identity, α- and β-tubulin have very similar three- dimensional structures. Further analysis revealed that the tubulins are members of the P-loop NTPase family and contain a nucleotide-binding site adjacent to the P-loop. Tubulins are present only in eukaryotes, although a prokaryotic homolog has been found. Sequence analysis identified a prokaryotic protein called FtsZ (for *f*ilamentous temperature-sensitive mutant *Z*) that is quite similar to the tubulins. The homology was confirmed when the structure was determined by x-ray crystallography. Interestingly, this protein participates in bacterial cell division, forming ring-shaped structures at the constriction that arises when a cell divides. These observations suggest that tubulins may have evolved from an ancient cell-division protein.

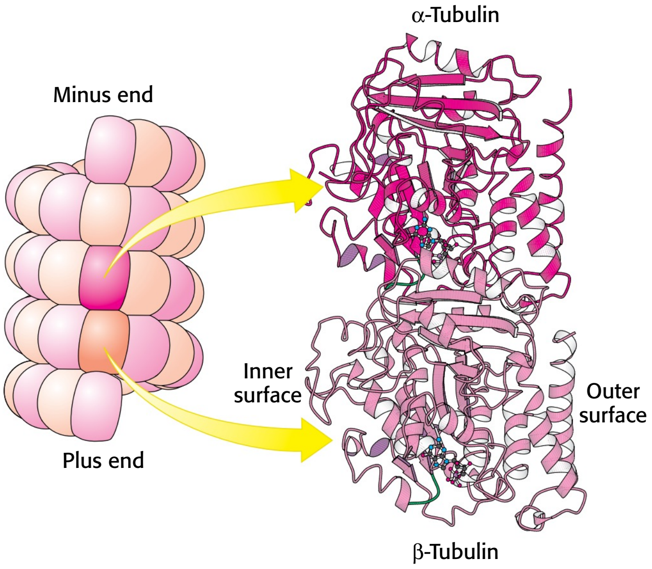
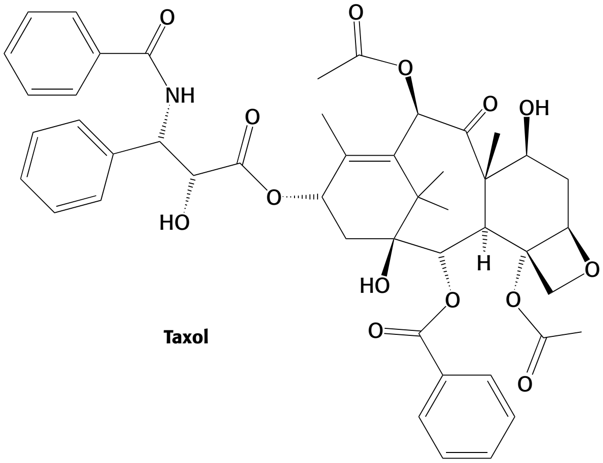


FIGURE 35.22 Tubulin. Microtubules can be viewed as an assembly of α-tubulin–β-tubulin dimers. The structures of α-tubulin and β-tubulin are quite similar. *Notice* that each includes a P-loop NTPase domain (shown in purple) and a bound guanine nucleotide. [Drawn from 1JFF.pdb.]

The continual lengthening and shortening of microtubules is essential to their role in cell division. *Taxol,* a compound isolated from the bark of the Pacific yew tree, was discovered through its ability to interfere with cell proliferation.



Taxol

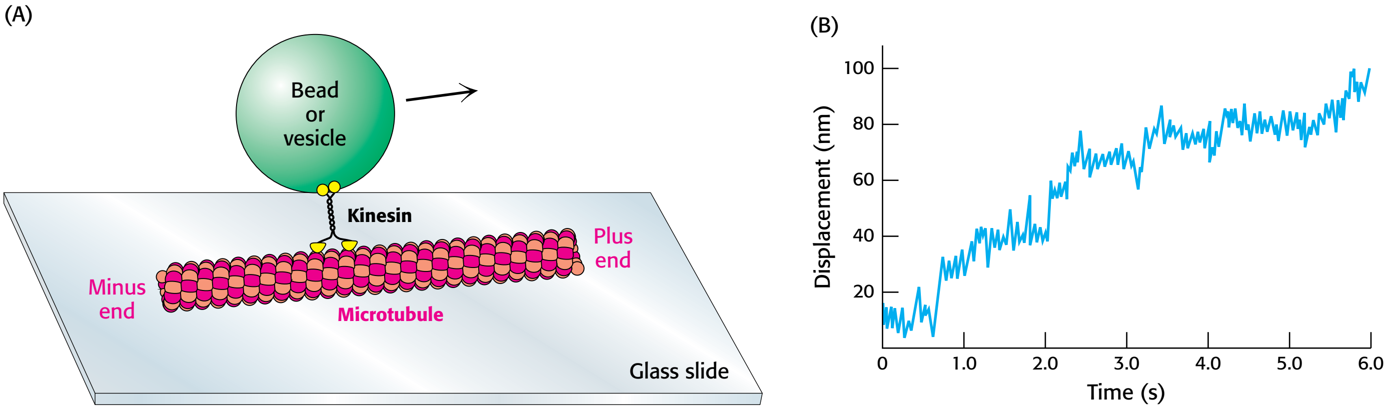
Taxol binds to microtubules and stabilizes the polymerized form. Taxol and its derivatives have been developed as anticancer agents because they preferentially affect rapidly dividing cells, such as those in tumors.

Kinesin motion is highly processive

Kinesins are motor proteins that move along microtubules. We have seen that muscle myosin moves along actin filaments by a process in which actin is released in each cycle; a myosin head group acting independently dissociates from actin after every power stroke. In contrast, when a kinesin molecule moves along a microtubule, the two head groups of the kinesin molecule operate in tandem: one binds, and then the next one does. A kinesin molecule may take many steps before both head groups are dissociated at the same time. In other words, the motion of kinesin is highly processive. Single-molecule measurements allow processive motion to be observed (Figure 35.23). A single kinesin molecule will typically take 100 or more steps toward the plus end of a microtubule in a period of seconds before the molecule becomes detached from the microtubule. These measurements also revealed that the average step size is approximately 80 Å, a value that corresponds to the distance between consecutive α- or β-tubulin subunits along each protofilament.

An additional fact is crucial to the development of a mechanism for kinesin motion—namely, that the addition of ATP strongly *increases* the affinity of kinesin for microtubules. This behavior stands in contrast with the behavior of myosin; ATP binding to myosin promotes its *dissociation* from actin. Do these differences imply that kinesin and myosin operate by completely different mechanisms? Indeed not. Kinesin-generated movement

FIGURE 35.23 Monitoring movements mediated by kinesin. (A) The movement of beads or vesicles, carried by individual kinesin dimers along a microtubule, can be directly observed. (B) A trace shows the displacement of a bead carried by a kinesin molecule. Multiple steps are taken in the 6-s interval. The average step size is about 8 nm (80 Å). [(B) Data from K. Svoboda et al., *Nature* 365:721–727, 1993.]



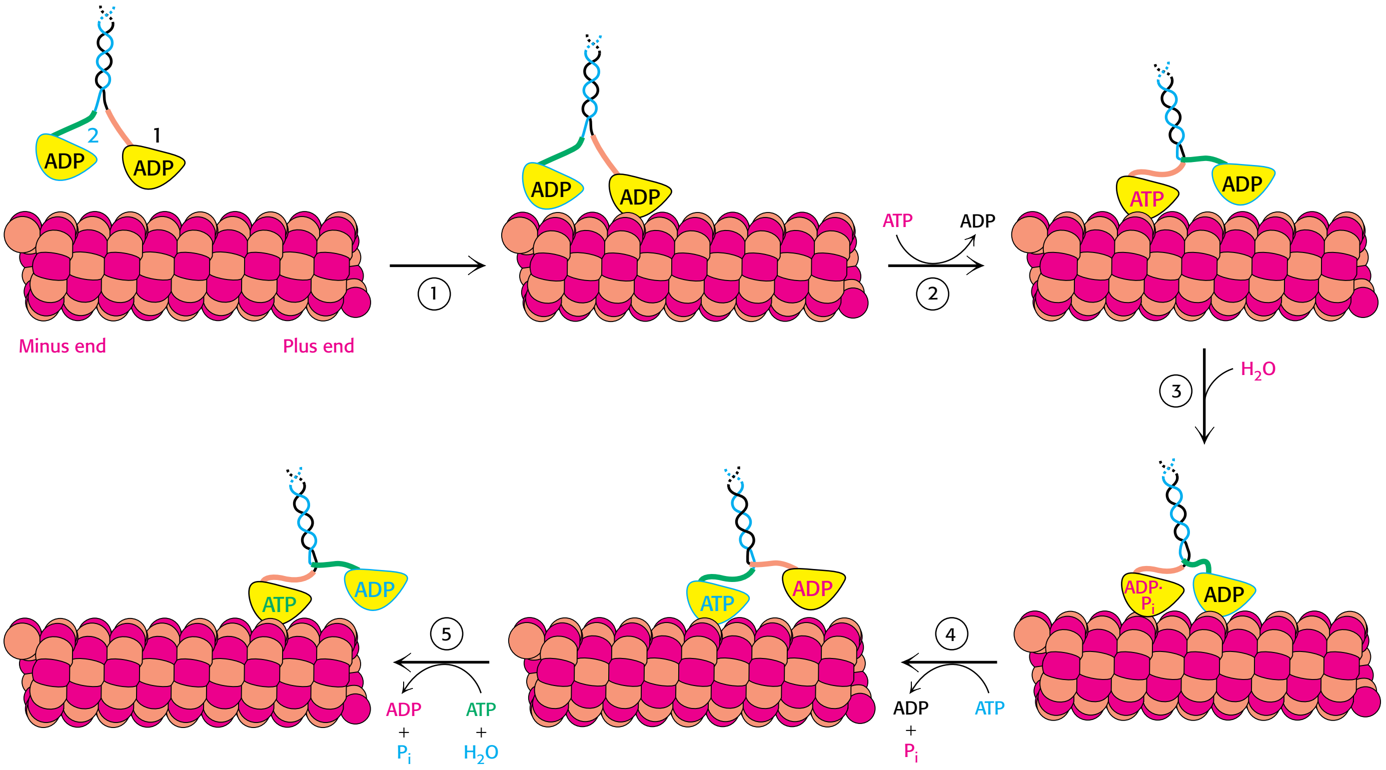


FIGURE 35.24 Kinesin moving along a microtubule. (1) One head of a two-headed kinesin molecule, initially with both heads in the ADP form, binds to a microtubule. (2) The release of ADP and the binding of ATP results in a conformational change that locks the head to the microtubule and pulls the neck linker (orange) to the head domain, throwing the second domain toward the plus end of the microtubule. (3) ATP undergoes hydrolysis while the second head interacts with the microtubule. (4) The exchange of ATP for ADP in the second head pulls the first head off the microtubule, releasing P_i and moving the first domain along the microtubule. (5) The cycle repeats, moving the kinesin dimer farther down the microtubule.

appears to proceed by a mechanism that is quite similar to that used by myosin (Figure 35.24). Let us begin with a two-headed kinesin molecule in its ADP form, dissociated from a microtubule. Recall that the neck linker binds the head domain when ATP is bound and is released when ADP is bound. The initial interaction of one of the head domains with a tubulin dimer on a microtubule stimulates the release of ADP from this head domain and the subsequent binding of ATP. The binding of ATP triggers a conformational change in the head domain that leads to two important events. First, the affinity of the head domain for the microtubule increases, essentially locking this head domain in place. Second, the neck linker binds to the head domain. This change, transmitted through the coiled-coil domain that connects the two kinesin monomers, repositions the other head domain. In its new position, the second head domain is close to a second tubulin dimer, 80 Å along the microtubule in the direction of the plus end. Meanwhile, the intrinsic ATPase activity of the first head domain hydrolyzes the ATP to ADP and P_i. When the second head domain binds to the microtubule, the first head releases ADP and binds ATP. Again, ATP binding favors a conformational change that pulls the first domain forward. This process can continue for many cycles until, by chance, both head domains are in the ADP form simultaneously and kinesin dissociates from the microtubule. Because of the relative rates of the component reactions, a simultaneous dissociation takes place approximately every 100 cycles. Thus, kinesin "walks" hand over hand along the microtubule for approximately 100 steps before both "hands" simultaneously let go. Although recent studies have led to a structural model for the dynein power stroke, the mechanism by which it works is unknown and an interesting research frontier.

Kinesin hydrolyzes ATP at a rate of approximately 80 molecules per second. Thus, given the step size of 80 Å per molecule of ATP, kinesin moves along a microtubule at a speed of 6400 Å per second. This rate is considerably lower than the maximum rate for myosin, which moves relative to actin at

80,000 Å per second. Recall, however, that myosin movement depends on the independent action of hundreds of different head domains working along the same actin filament, whereas the movement of kinesin is driven by the processive action of kinesin head groups working in pairs. Muscle myosin evolved to maximize the speed of the motion, whereas kinesin functions to achieve steady, but slower, transport in one direction along a filament.

35.4 A Rotary Motor Drives Bacterial Motion

In 1 s, a motile bacterium can move approximately 25 μm, or about 10 body lengths. A human being sprinting at a proportional rate would complete the 100-meter dash in slightly more than 5 s. The motors that power this impressive motion are strikingly different from the eukaryotic motors that we have seen so far. In the bacterial motor, an element spins around a central axis rather than moving along a polymeric track. The direction of rotation can change rapidly, a feature that is central to chemotaxis, the process by which bacteria swim preferentially toward an increasing concentration of certain useful compounds and away from potentially harmful ones. One type of flagellar motor, powered by a Na^+ gradient, turns at a rate of 200,000 revolutions per minute.

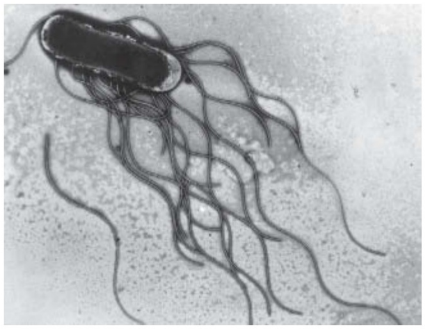


FIGURE 35.25 Bacterial flagella. Electron micrograph of *Salmonella typhimurium* shows flagella directed backwards, where they can form a bundle under appropriate circumstances. [Science Source.]

Bacteria swim by rotating their flagella

Bacteria such as *Escherichia coli* and *Salmonella typhimurium* swim by rotating flagella that lie on their surfaces (Figure 35.25). When the flagella rotate in a counterclockwise direction (viewed from outside a bacterium), the separate flagella form a bundle that very efficiently propels the bacterium through solution.

Bacterial flagella are polymers approximately 15 nm in diameter and as much as 15 μm in length, composed of 53-kDa subunits of a protein called *flagellin* (Figure 35.26). These subunits associate into a helical structure that has 5.5 subunits per turn, giving the appearance of 11 protofilaments. Each flagellum has a hollow core. Remarkably, flagella form not by growing at the base adjacent to the cell body but, instead, by the addition of new subunits that pass through the hollow core and add to the free end. Each flagellum is intrinsically twisted in a left-handed sense. At its base, each flagellum has a rotory motor.

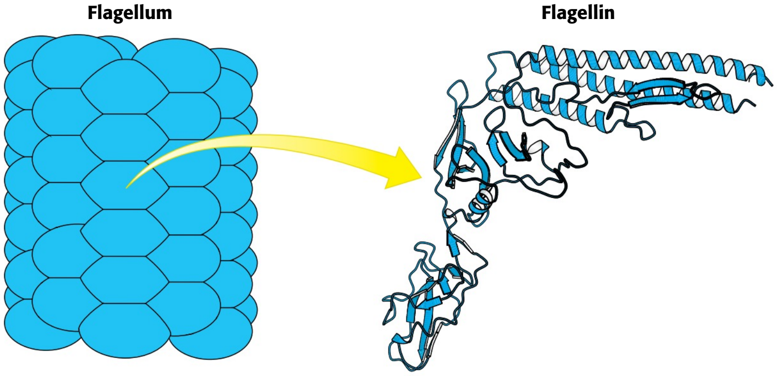


FIGURE 35.26 Structure of flagellin. A bacterial flagellum is a helical polymer of the protein flagellin. *Notice* that each subunit corresponds to a bent structure with a relatively flat surface facing the hollow core of the flagellum. [Drawn from 1IO1.pdb.]

Proton flow drives bacterial flagellar rotation

Early experiments by Julius Adler demonstrated that ATP is *not* required for flagellar motion. What powers these rotary motors? The necessary free energy is derived from the proton gradient that exists across the plasma membrane. The flagellar motor is quite complex, containing as many as 40 distinct proteins (Figure 35.27). Five components particularly crucial to motor function have been identified through genetic studies. MotA is a membrane protein that appears to have four transmembrane helices as well as a cytoplasmic domain. MotB is another membrane protein with a single trans-membrane helix and a large periplasmic domain. Approximately 11 *MotA–MotB pairs* form a ring around the base of the flagellum. The proteins *FliG*, *FliM*, and *FliN* are part of a disc-like structure called the MS (membrane and supramembrane) ring, with approximately 30 FliG subunits coming together to form the ring. The three-dimensional structure of the carboxyl-terminal half of FliG reveals a wedge-shaped domain with a set of charged amino acids, conserved among many species, lying along the thick edge of the wedge (Figure 35.28).

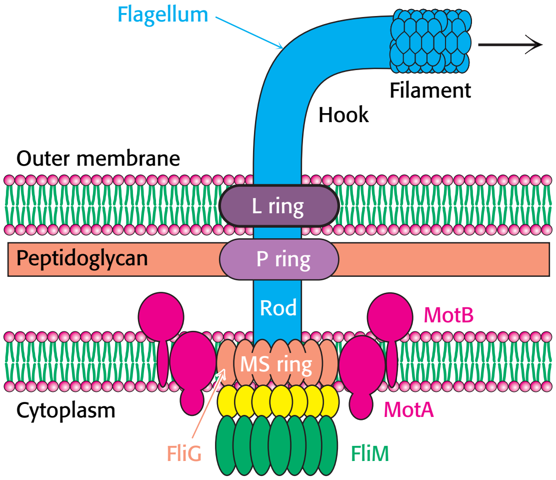


FIGURE 35.27 Flagellar motor. A schematic view of the flagellar motor, a complex structure containing as many as 40 distinct types of protein. The approximate positions of the proteins MotA and MotB (red), FliG (orange), FliN (yellow), and FliM (green) are shown.

The MotA–MotB pair and FliG combine to create a proton channel that drives the rotation of the flagellum. How can proton flow across a membrane drive mechanical rotation? We have seen such a process earlier in regard to ATP synthase (Section 18.4). Recall that the key to driving the rotation of the γ subunit of ATP synthase is the **a** subunit of the F_0 fragment. This subunit appears to have two half-channels; protons can move across the membrane only by moving into the half-channel from the side of the membrane with the higher local proton concentration, binding to a disc-like structure formed by the **c** subunits, riding on this structure as it rotates to the opening of the other half-channel, and exiting to the side with the lower local proton concentration. Could a similar mechanism apply to flagellar rotation? Indeed, such a mechanism was first proposed by Howard Berg to explain flagellar rotation before the rotary mechanism of ATP synthase was elucidated. Each MotA–MotB pair is conjectured to form a structure that has two half-channels; FliG serves as the rotating proton carrier, perhaps with the participation of some of the charged residues identified in crystallographic studies (Figure 35.29). In this scenario, a proton from the periplasmic space passes into the outer half-channel and is transferred to an FliG subunit. The MS ring rotates, rotating the flagellum with it and allowing the proton to pass into the inner half-channel and into the cell. Ongoing structural and mutagenesis studies are testing and refining this hypothesis.

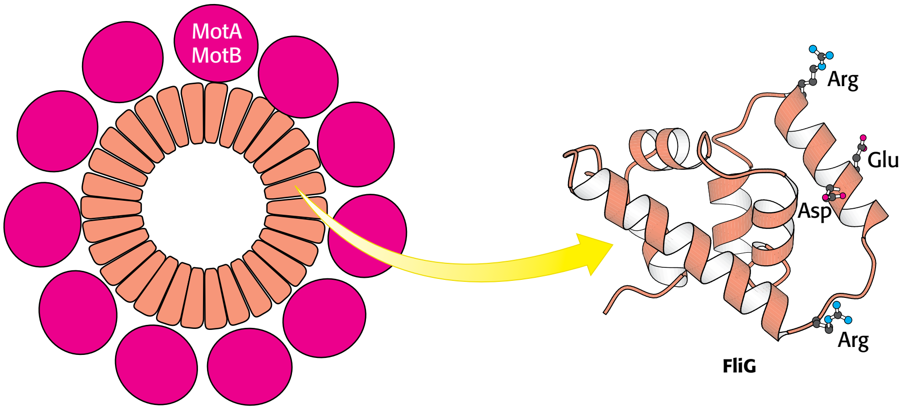


FIGURE 35.28 Flagellar motor components. Approximately 30 subunits of FliG assemble to form part of the MS ring. The ring is surrounded by approximately 11 structures consisting of MotA and MotB. *Notice* that the carboxyl-terminal domain of FliG includes a ridge lined with charged residues that may participate in proton transport. [Drawn from 1QC7.pdb.]

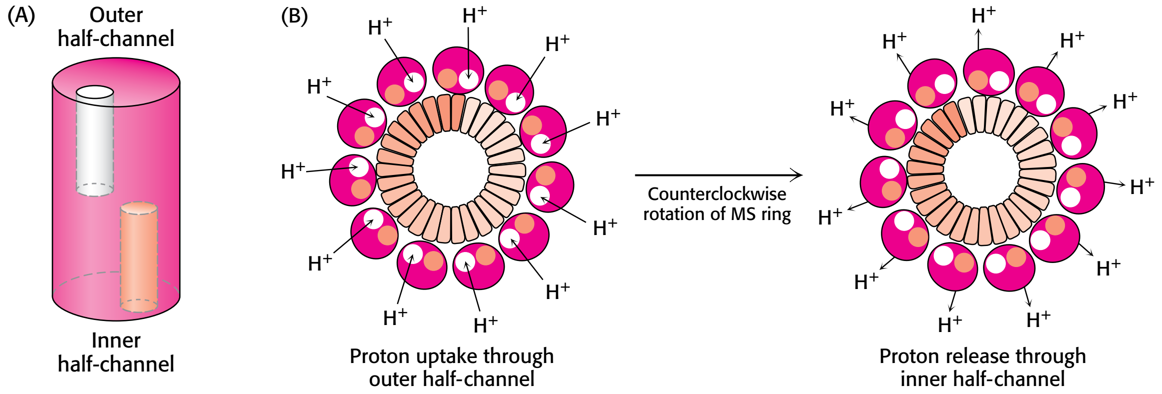

(A)
Outer half-channel
Inner half-channel
(B)
H+
Counterclockwise rotation of MS ring
Proton uptake through outer half-channel
Proton release through inner half-channel

FIGURE 35.29 Proton-transport-coupled rotation of the flagellum. (A) MotA–MotB may form a structure having two half-channels. (B) One model for the mechanism of coupling rotation to a proton gradient requires protons to be taken up into the outer half-channel and transferred to the MS ring. The MS ring rotates in a counterclockwise direction, and the protons are released into the inner half-channel. The flagellum is linked to the MS ring and so the flagellum rotates as well.

Bacterial chemotaxis depends on reversal of the direction of flagellar rotation

Many species of bacteria respond to changes in their environments by adjusting their swimming behavior. Examination of the paths taken is highly revealing (Figure 35.30). The bacteria swim in one direction for some length of time (typically about a second), tumble briefly, and then set off in a new direction. The tumbling is caused by a brief reversal in the direction of the flagellar motor. When the flagella rotate counterclockwise, the helical filaments form a coherent bundle favored by the intrinsic shape of each filament, and the bacterium swims smoothly. When the rotation reverses, the bundle flies apart because the screw sense of the helical flagella does not match the direction of rotation. Each flagellum then pulls in a different direction and the cell tumbles.

In the presence of a gradient of certain substances such as glucose, bacteria swim preferentially toward the direction of the higher concentration of the substance. Such compounds are referred to as *chemoattractants*. Bacteria also swim preferentially away from potentially harmful compounds such as phenol, a *chemorepellant*. The process of moving in specific directions in response to environmental cues is called *chemotaxis*. In the presence of a gradient of a chemoattractant, bacteria swim for longer periods of time without tumbling when moving toward higher concentrations of the chemoattractant. In contrast, they tumble more frequently when moving toward lower concentrations of the chemoattractant. This behavior is reversed for chemorepellants. The result of these actions is a *biased random walk* that facilitates net motion toward conditions more favorable to the bacterium.

Chemotaxis depends on a signaling pathway that terminates at the flagellar motor. The signaling pathway begins with the binding of molecules to receptors in the plasma membrane (Figure 35.31). In their *unoccupied* forms, these receptors initiate a pathway leading eventually to the phosphorylation of a specific aspartate residue on a soluble protein called *CheY*. In its phosphorylated form, CheY binds to the base on the flagellar motor. When bound to phosphorylated CheY, the flagellar motor rotates in a clockwise rather than a counterclockwise direction, causing tumbling.

The binding of a chemoattractant to a surface receptor blocks the signaling pathway leading to CheY phosphorylation. Phosphorylated CheY spontaneously hydrolyzes and releases its phosphoryl group in a process accelerated by another protein, CheZ. The concentration of phosphorylated CheY drops, and the flagella are less likely to rotate in a clockwise direction. Under these conditions, bacteria swim smoothly without tumbling. Thus, the reversible rotary flagellar motor and a phosphorylation-based signaling pathway work together to generate an effective means for responding to environmental conditions.

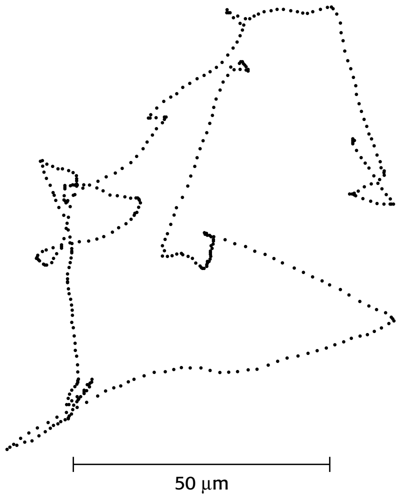


FIGURE 35.30 Charting a course. This projection of the track of an *E. coli* bacterium was obtained with a microscope that automatically follows bacterial motion in three dimensions. The points show the locations of the bacterium at 80-ms intervals. [Information from H. C. Berg, *Nature* 254:389–392, 1975.]

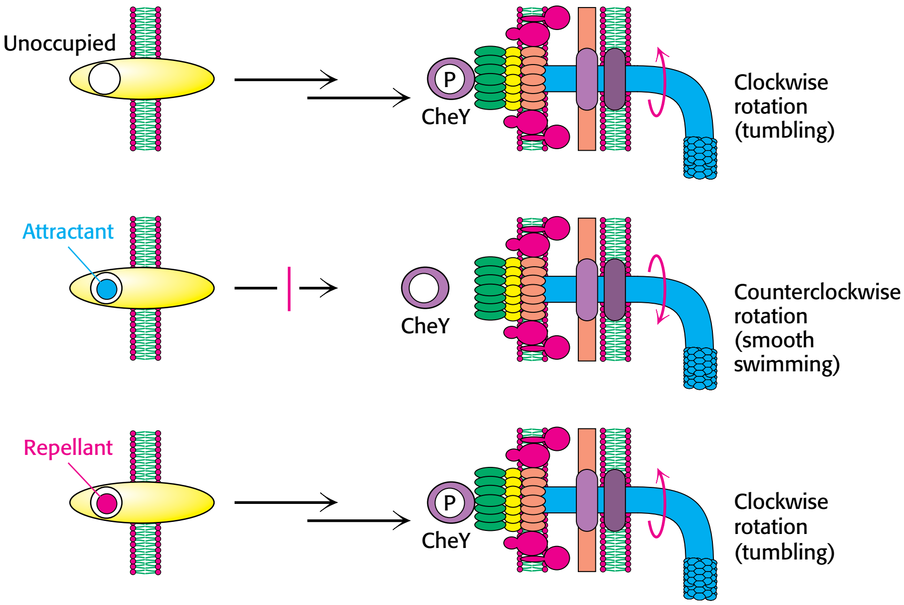

Unoccupied
P
CheY
Clockwise rotation (tumbling)
Attractant
CheY
Counterclockwise rotation (smooth swimming)
Repellant
P
CheY
Clockwise rotation (tumbling)

FIGURE 35.31 Chemotaxis signaling pathway. Receptors in the plasma membrane initiate a signaling pathway leading to the phosphorylation of the CheY protein. Phosphorylated CheY binds to the flagellar motor and favors clockwise rotation. When an attractant binds to the receptor, this pathway is blocked, and counterclockwise flagellar rotation and, hence, smooth swimming result. When a repellant binds, the pathway is stimulated, leading to an increased concentration of phosphorylated CheY and, hence, more-frequent clockwise rotation and tumbling.

Bacteria sense spatial gradients of chemoattractants by measurements separated in time. A bacterium sets off in a random direction and, if the concentration of the chemoattractant has increased after the bacterium has been swimming for a period of time, the likelihood of tumbling decreases and the bacterium continues in roughly the same direction. If the concentration has decreased, the tumbling frequency increases and the bacterium tests other random directions. The success of this mechanism once again reveals the power of evolutionary problem solving: many possible solutions are tried at random, and those that are beneficial are selected and exploited.

SUMMARY

35.1 Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily

Eukaryotic cells contain three families of molecular-motor proteins: myosins, kinesins, and dyneins. These proteins move along tracks defined by the actin and microtubule cytoskeletons of eukaryotic cells, contributing to cell and organismal movement and to the intracellular transport of proteins, vesicles, and organelles. Despite considerable differences in size and a lack of similarity detectable at the level of amino acid sequence, these proteins are homologous, containing core structures of the P-loop NTPase family. The ability of these core structures to change conformations in response to nucleoside triphosphate binding and hydrolysis is key to molecular-motor function. Motor proteins consist of motor domains attached to extended structures that serve to amplify the conformational changes in the core domains and to link the core domains to one another or to other structures.

35.2 Myosins Move Along Actin Filaments

The motile structure of muscle consists of a complex of myosin and actin, along with accessory proteins. Actin, a highly abundant 42-kDa protein, polymerizes to form long filaments. Each actin monomer can bind either ATP or ADP. A myosin motor domain moves along actin filaments in a cyclic manner, beginning with myosin free of bound nucleotides bound to actin: (1) ATP binds to myosin and the myosin is released from actin; (2) a reversible conformational change associated with the hydrolysis of ATP while still bound to myosin leads to a large motion of a lever arm that extends from the motor domain; (3) myosin with bound ADP and P_i binds to actin; (4) P_i is released from myosin, resulting in resetting the position of the lever arm and moving actin relative to myosin; and (5) the release of ADP returns the motor domain to its initial state. The length of the lever arm determines the size of the step taken along actin in each cycle. The ability to monitor single molecular-motor proteins has provided key tests for hypotheses concerning motor function. Muscle contraction entails the rapid sliding of thin filaments, composed of actin, relative to thick filaments, composed of myosin. The thick filaments consist of multiple myosin molecules bound together. Each myosin molecule has two heads that can bind to actin and move it relative to myosin, driven by the hydrolysis of ATP by myosin. Muscle contraction is regulated by tropomyosin and the troponin complex. These proteins prevent actin and myosin from productively interacting until an increase in calcium concentration associated with a nerve impulse results in calcium-induced changes in troponin and tropomyosin.

35.3 Kinesin and Dynein Move Along Microtubules

Kinesin and dynein move along microtubules rather than actin. Microtubules are polymeric structures composed of α- and β-tubulin, two very similar guanine-nucleotide-binding proteins. Each microtubule comprises 13 protofilaments with alternating α- and β-tubulin subunits. Kinesins move along microtubules by a mechanism quite similar to that used by myosin to move along actin but with several important differences. First, ATP binding to kinesin favors motor-domain binding rather than dissociation. Second, the power stroke is triggered by the binding of ATP rather than the release of P_i. Finally, kinesin motion is processive. The two heads of a kinesin dimer work together, taking turns binding and releasing the microtubule, and many steps are taken along a microtubule before both heads dissociate. Most kinesins move toward the plus end of microtubules.

35.4 A Rotary Motor Drives Bacterial Motion

Many motile bacteria use rotating flagella to propel themselves. When rotating counterclockwise, multiple flagella on the surface of a bacterium come together to form a bundle that effectively propels the bacterium through solution. A proton gradient across the plasma membrane, rather than ATP hydrolysis, powers the flagellar motor. The mechanism for coupling transmembrane proton transport to macromolecular rotation appears to be similar to that used by ATP synthase. When rotating clockwise, the flagella fly apart and the bacterium tumbles. Bacteria swim preferentially toward chemoattractants in a process called chemotaxis. When bacteria are swimming in the direction of an increasing concentration of a chemoattractant, clockwise flagellar motion predominates and tumbling is suppressed, leading to a biased random walk in the direction of increasing chemoattractant concentration.

KEY TERMS

cytoskeleton (p. 1011)
myosin (p. 1012)
kinesin (p. 1012)
dynein (p. 1012)
S1 fragment (p. 1013)
conventional kinesin (p. 1014)
lever arm (p. 1015)
relay helix (p. 1016)
neck linker (p. 1016)
actin (p. 1016)
G-actin (p. 1017)
F-actin (p. 1017)

critical concentration (p. 1018)
optical trap (p. 1019)
power stroke (p. 1019)
myofibril (p. 1020)
sarcomere (p. 1020)
thick filaments (p. 1020)
thin filaments (p. 1020)
tropomyosin (p. 1020)
troponin complex (p. 1020)
microtubule (p. 1022)
axoneme (p. 1023)
tubulin (p. 1023)
dynamic instability (p. 1023)
flagellin (p. 1026)
MotA–MotB pair (p. 1027)
FliG (p. 1027)
chemoattractant (p. 1028)
chemorepellant (p. 1028)
chemotaxis (p. 1028)
CheY (p. 1028)

PROBLEMS

1. *Diverse motors.* Skeletal muscle, eukaryotic cilia, and bacterial flagella use different strategies for the conversion of free energy into coherent motion. Compare and contrast these motility systems with respect to (a) the free-energy source and (b) the number of essential components and their identity.

2. *You call that slow?* At maximum speed, a kinesin molecule moves at a rate of 6400 Å per second. Given the dimensions of the motor region of a kinesin dimer of approximately 80 Å, calculate its speed in "body lengths" per second. To what speed does this body-length speed correspond for an automobile 10 feet long?

3. *Heavy lifting.* A single myosin motor domain can generate a force of approximately 4 piconewtons (4 pN). How many times its "body weight" can a myosin motor domain lift? Note that 1 newton = 0.22 pounds (100 g). Assume a molecular mass of 100 kDa for the motor domain.

4. *Compare and contrast.* Describe two similarities and two differences between actin filaments and microtubules.

5. *Lighten up.* What is the primary role of the light chains in myosin? In kinesin?

6. *Rigor mortis.* Propose an explanation for the fact that the body stiffens after death.

7. *Now you see it, now you don't.* Under certain stable concentration conditions, actin monomers in their ATP form will polymerize to form filaments that disperse again into free actin monomers over time. Explain.

8. *Helicases as motors.* Helicases can use single-stranded DNA as tracks. Consider a helicase that moves one base in the $3' \rightarrow 5'$ direction in each cycle. Assuming that the helicase can hydrolyze ATP at a rate of 50 molecules per second in the presence of a single-stranded DNA template, calculate the velocity of the helicase in micrometers per second. How does this velocity compare with that of kinesin?

9. *New moves.* When bacteria such as *E. coli* are starved to a sufficient extent, they become nonmotile. However, when such bacteria are placed in an acidic solution, they resume swimming. Explain.

10. *Going straight.* Suppose you measure the mean distance that an *E. coli* bacterium moves along a straight path before it tumbles over a period of time. Would you expect this distance to change in the presence of a gradient of a chemoattractant? Would it be longer or shorter?

11. *Hauling a load.* Consider the action of a single kinesin molecule in moving a vesicle along a microtubule track. The force required to drag a spherical particle of radius a at a velocity v in a medium having a viscosity η *is*

$$F = 6\pi\eta a v$$

Suppose that a 2-μm-diameter bead is carried at a velocity of 0.6 $\mu m\ s^{-1}$ in an aqueous medium ($\eta = 0.01$ poise $= 0.01\ g\ cm^{-1}\ s^{-1}$).

(a) What is the magnitude of the force exerted by the kinesin molecule? Express the value in dynes ($1\ dyne = 1\ g\ cm\ s^{-2}$).

(b) How much work is performed in 1 s? Express the value in ergs (1 erg = 1 dyne cm).

(c) A kinesin motor hydrolyzes approximately 80 molecules of ATP per second. What is the energy associated with the hydrolysis of this much ATP in ergs? Compare this value with the actual work performed.

12. *Unusual strides.* A publication describes a kinesin molecule that is claimed to move along microtubules with a step size of 6 nm. You are skeptical. Why?

13. *The sound of one hand clapping.* KIF1A is a motor protein that moves toward the plus end of microtubules *as a monomer.* KIF1A has only a single motor domain. What additional structural elements would you expect to find in the KIF1A structure?

14. *Building blocks.* Actin filaments, microtubules, and bacterial flagella are all built from small subunits. Describe three advantages of assembling long filamentous structures from subunits rather than from single, long proteins.

Mechanism Problem

15. *Backward rotation.* On the basis of the proposed structure in Figure 35.29 for the bacterial flagellar motor, suggest

a pathway for transmembrane proton flow when the flagellar motor is rotating clockwise rather than counterclockwise.

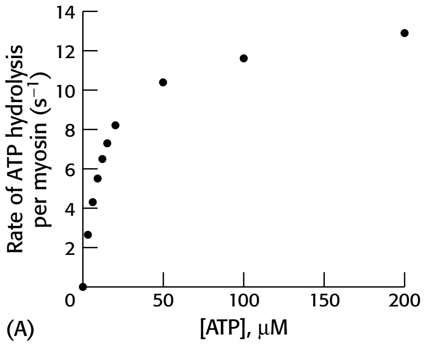
Chapter Integration Problem

16. *Smooth muscle.* Smooth muscle, in contrast with skeletal muscle, is not regulated by a tropomyosin–troponin mechanism. Instead, vertebrate smooth-muscle contraction is controlled by the degree of phosphorylation of its light chains. Phosphorylation induces contraction, and dephosphorylation leads to relaxation. Like that of skeletal muscle, smooth-muscle contraction is triggered by an increase in the cytoplasmic calcium ion level. Propose a mechanism for this action of calcium ion on the basis of your knowledge of other signal-transduction processes.

Data Interpretation Problem

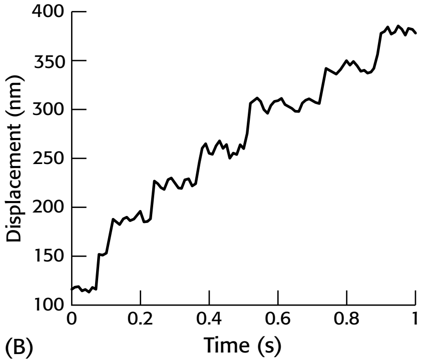
17. *Myosin V.* An abundant myosin-family member, myosin V is isolated from brain tissue. This myosin has a number of unusual properties. First, on the basis of its amino acid sequence, each heavy chain has six tandem binding sites for calmodulin-like light chains. Second, it forms dimers but not higher-order oligomers. Finally, unlike almost all other myosin-family members, myosin V is highly processive.

The rate of ATP hydrolysis by myosin has been examined as a function of ATP concentration, as shown in graph A.



(a) Estimate the values of k_{cat} and K_M for ATP.

With the use of optical-trap measurements, the motion of single myosin V dimers could be followed, as shown in graph B.



[Based on M. Rief et al., *Proc. Natl. Acad. Sci. U. S. A.* 97:9482–9486, 2000.]

(b) Estimate the step size for myosin V.

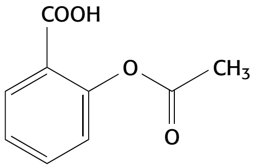
The rate of ADP release from myosin V is found to be approximately 13 molecules per second.

(c) Combine the observations about the amino acid sequence of myosin, the observed step size, and the kinetics results to propose a mechanism for the processive motion of myosin V.

CHAPTER

36

Drug Development



Many drugs are based on natural products. Aspirin (above) is a molecule related to a compound isolated from willow bark (near left). Extracts of willow bark had long been known to have medicinal properties. The active compound was isolated, modified, and, beginning in 1899, packaged for consumers (far left). [(Far left) AP Images/Bayer. (Near left) H. Mark Weidman Photography/Alamy.]

OUTLINE

36.1 The Development of Drugs Presents Huge Challenges

36.2 Drug Candidates Can Be Discovered by Serendipity, Screening, or Design

36.3 Analyses of Genomes Hold Great Promise for Drug Discovery

36.4 The Clinical Development of Drugs Proceeds Through Several Phases

The development of drugs is among the most important interfaces between biochemistry and medicine. Throughout the previous chapters in this book, we have encountered a number of types of proteins—enzymes, receptors, and transporters—that serve as targets for a majority of the drugs in clinical use (Figure 36.1). In most cases, drugs act by binding to these proteins and inhibiting or otherwise modulating their activities. Thus, knowledge of these molecules and the pathways in which they participate is crucial to drug development. An effective drug is much more than a potent modulator of its target, however. Drugs must be readily administered to patients, preferably as small tablets taken orally, and must survive within the body long enough to reach their targets. Furthermore, to prevent unwanted physiological effects, drugs must not modulate the properties of biomolecules other than the intended targets. These requirements tremendously limit the number of compounds that have the potential to be clinically useful drugs.

Drugs have been discovered by two fundamentally distinct approaches (Figure 36.2). The first approach identifies a substance that has a desirable physiological consequence when administered to a human being, to an appropriate animal, or to cells. Such substances can be discovered by serendipity, by the fractionation of plants or other materials known to have

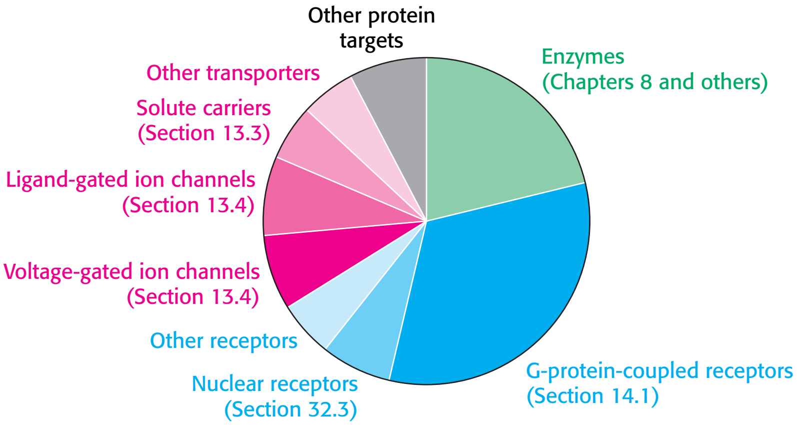


FIGURE 36.1 The targets of drugs in current use. In this pie chart, currently-used drugs that modulate human proteins are sorted by target type: enzymes in green, receptors in blue, transporters in red, and other targets in gray. [Data from M. Rask-Andersen et al., *Nat. Rev. Drug Discov.* 10:579–590, 2011]

medicinal properties, or by screening natural products or other "libraries" of compounds. In this approach, a biological effect is known before the molecular target is identified. The mode of action of the substance is only later identified after substantial additional work. The second approach begins with a known molecular target. Compounds are sought, either by screening or by designing molecules with desired properties, that bind to the target molecule and modulate its properties. After such compounds are available, scientists can explore their effects on appropriate cells or organisms. Many unexpected results may be encountered in this process as the complexity of biological systems reveals itself.

In this chapter, we explore the science of pharmacology. We examine a number of case histories that illustrate drug development—including many of its concepts, methods, and challenges. We then see how the concepts and tools from genomics are influencing approaches to drug discovery. The chapter ends with a summary of the clinical trial phases required for drug development.

Pharmacology
The science that deals with the discovery, chemistry, composition, identification, biological and physiological effects, uses, and manufacture of drugs.

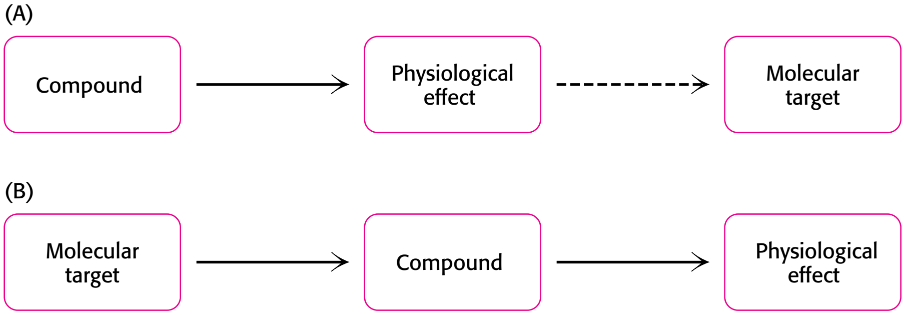


FIGURE 36.2 Two paths to drug discovery. (A) A compound is discovered to have a desirable physiological effect. The molecular target can be identified in a separate step as needed. (B) A molecular target is selected first. Drug candidates that bind to the target are identified and then examined for their physiological effects.

36.1 The Development of Drugs Presents Huge Challenges

Many compounds have significant effects when taken into the body, but only a very small fraction of them have the potential to be useful drugs. A foreign compound, not adapted to its role in the cell through long evolution, must have a range of special properties to function effectively without causing serious harm. Let us consider some of the challenges faced by drug developers.

Drug candidates must be potent and selective modulators of their targets

Most drugs bind to specific proteins, usually receptors or enzymes, within the body. To be effective, a drug should bind a sufficient number of its target proteins when taken at a reasonable dose. One factor in determining drug effectiveness is the strength of the interaction between the drug and its target. A molecule that binds to some target protein is often referred to as a *ligand.* A ligand-binding curve is shown in Figure 36.3. Ligand molecules occupy progressively more target binding sites as ligand concentration increases until essentially all of the available sites are occupied. The tendency of a ligand to bind to its target is measured by the *dissociation constant*, K_d, defined by the expression

$$K_d = [R][L]/[RL]$$

where [R] is the concentration of the free receptor, [L] is the concentration of the free ligand, and [RL] is the concentration of the receptor–ligand complex. The dissociation constant is a measure of the strength of the interaction between the drug candidate and the target; the lower the value, the stronger the interaction. The concentration of free ligand at which one-half of the binding sites are occupied equals the dissociation constant, as long as the concentration of binding sites is substantially less than the dissociation constant.

In many cases, biological assays using living cells or tissues (rather than direct enzyme or binding assays) are employed to examine the potency of drug candidates. For example, the fraction of bacteria killed by a drug might indicate the potency of a potential antibiotic. In these cases, values such as EC_{50} are used. EC_{50} is the concentration of the drug candidate required to elicit 50% of the maximal biological response (Figure 36.4). Similarly, EC_{90} is the concentration required to achieve 90% of the maximal response. In the example of an antibiotic, EC_{90} would be the concentration required to kill 90% of bacteria exposed to the drug. For drug candidates that are inhibitors, the corresponding terms IC_{50} and IC_{90} are often used to describe the concentrations of the inhibitor required to reduce a response by 50% or 90%, relative to its value in the absence of inhibitor, respectively.

Values such as the IC_{50} and EC_{50} are measures of the *potency* of a drug candidate in modulating the activity of the desired biological target. To prevent unwanted effects, often called *side effects,* ideal drug candidates should also be *selective*. That is, they should not bind biomolecules other than the target to any significant extent. Developing such a drug can be quite challenging, particularly if the drug target is a member of a large family of evolutionarily related proteins. The degree of selectivity can be described in terms of the ratio of the K_d values for the binding of the drug candidate to any other molecules to the K_d value for the binding of the drug candidate to the desired target.

Under physiological conditions, achieving sufficient binding of a target site by a drug can be quite challenging. Most drug targets also bind ligands normally present in tissues; often, the drug and these ligands will compete for binding sites on the target. We encountered this scenario when we considered competitive inhibitors in Chapter 8. Suppose that the drug target is an enzyme and the drug candidate is competitive with respect to the enzyme's natural substrate. The concentration of the drug candidate necessary to inhibit the enzyme will depend on the physiological concentration of the normal substrate (Figure 36.5). Biochemists Yung-Chi Cheng and William Prusoff

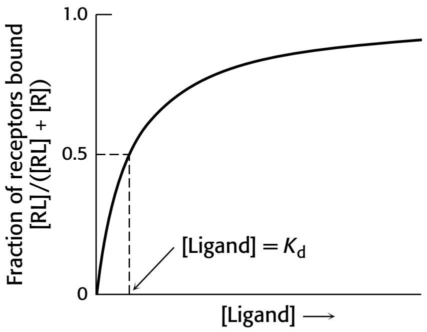


FIGURE 36.3 Ligand binding. The titration of a receptor, R, with a ligand, L, results in the formation of the complex RL. In uncomplicated cases, the binding reaction follows a simple saturation curve. Half of the receptors are bound to ligand when the ligand concentration equals the dissociation constant, K_d, for the RL complex.

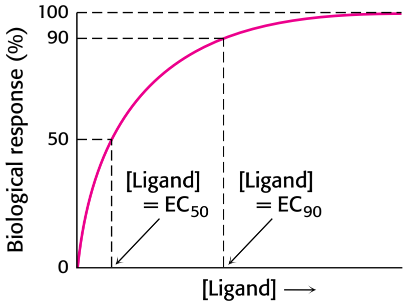


FIGURE 36.4 Effective concentrations. The concentration of a ligand required to elicit a biological response can be quantified in terms of EC_{50}, the concentration required to give 50% of the maximum response, and EC_{90}, the concentration required to give 90% of the maximum response.

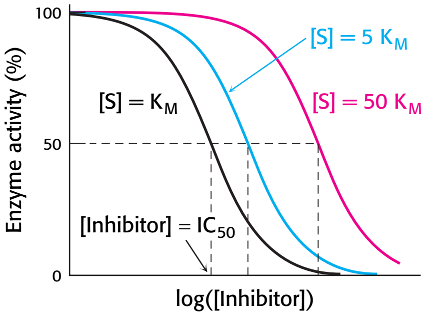


FIGURE 36.5 Inhibitors compete with substrates for enzyme active sites. The measured IC_{50} of a competitive inhibitor to its target enzyme depends on the concentration of substrate present.

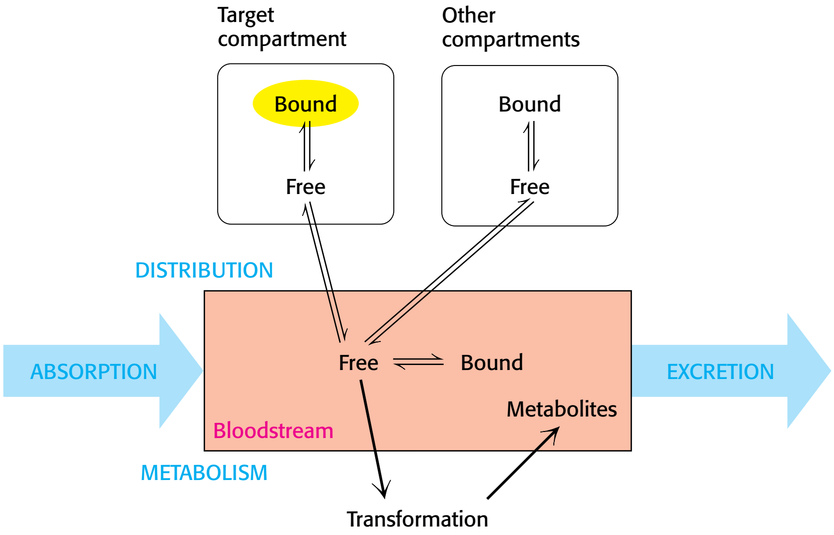


FIGURE 36.6 Absorption, distribution, metabolism, and excretion (ADME). The concentration of a compound at its target site (yellow) is affected by the extents and rates of absorption, distribution, metabolism, and excretion.

described the relation between the IC_{50} of an enzyme inhibitor and its *inhibition constant* K_i (analogous to the dissociation constant, K_d, of a ligand):

$$\mathrm{IC}_{50} = K_i\,(1 + [\mathrm{S}]/K_M)$$

This relation, referred to as the *Cheng–Prusoff equation*, demonstrates that the IC_{50} of a competitive inhibitor will depend on the concentration and the Michaelis constant (K_M) for the substrate S. The higher the concentration of the natural substrate, the higher the concentration of drug needed to inhibit the enzyme.

Drugs must have suitable properties to reach their targets

Thus far, we have focused on the ability of molecules to act on specific target molecules. However, an effective drug must also have other characteristics. It must be easily administered and must reach its target at sufficient concentration to be effective. After a drug molecule has entered the body, it is acted upon by several processes—absorption, distribution, metabolism, and excretion—that will determine the effective concentration of this molecule over time (Figure 36.6). A drug's response to these processes is referred to as its *ADME* (pronounced "add-me") properties.

Administration and absorption. Ideally, a drug can be taken orally as a small tablet. An orally administered active compound must be able to survive the acidic conditions in the gut and then be absorbed through the intestinal epithelium. Thus, the compound must be able to pass through cell membranes at a significant rate. Larger molecules such as proteins cannot be administered orally, because they often cannot survive the acidic conditions in the stomach and, if they do, are not readily absorbed. Even many small molecules are not absorbed well; they may be too polar and not pass through cell membranes easily, for example. The ability to be absorbed is often quantified in terms of the *oral bioavailability*. This quantity is defined as the ratio of the peak concentration of a compound given orally to the peak concentration of the same dose injected directly into the bloodstream. Bioavailability can vary considerably from species to species, and so results from animal studies may be difficult to apply to human beings. Despite this variability, some useful generalizations have been made. One powerful set consists of *Lipinski's rules.*

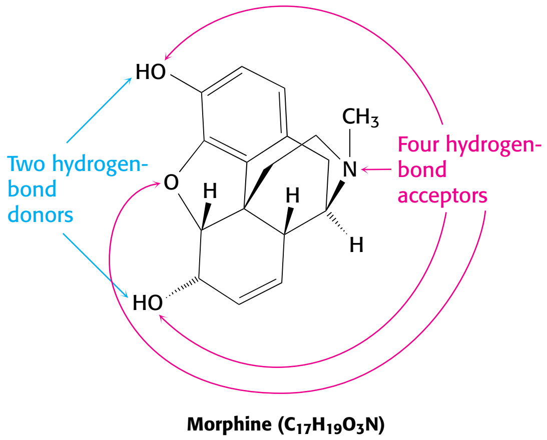
Lipinski's rules tell us that poor absorption is likely when

1. the molecular weight is greater than 500.
2. the number of hydrogen-bond donors is greater than 5.

3. the number of hydrogen-bond acceptors is greater than 10.
4. the partition coefficient [measured as log(P)] is greater than 5.

The partition coefficient is a way to measure the tendency of a molecule to dissolve in membranes, which correlates with its ability to dissolve in organic solvents. It is determined by allowing a compound to equilibrate between water and an organic phase, *n*-octanol. The log(P) value is defined as $\log_{10}$ of the ratio of the concentration of a compound in *n*-octanol to the concentration of the compound in water. For example, if the concentration of the compound in the *n*-octanol phase is 100 times that in the aqueous phase, then log(P) is 2. Although the ability of a drug to partition in organic solvents is ideal, because it implies that the compound can penetrate membranes, a log(P) value that is too high suggests that the molecule may be poorly soluble in an aqueous environment.

Morphine, for example, satisfies all of Lipinski's rules and has moderate bioavailability (Figure 36.7). A drug that violates one or more of these rules may still have satisfactory bioavailability. Nonetheless, these rules serve as guiding principles for evaluating new drug candidates.



Molecular weight = 285

log(P) = 1.27

FIGURE 36.7 Lipinski's rules applied to morphine. Morphine satisfies all of Lipinski's rules and has an oral bioavailability in human beings of 33%.

Distribution. Compounds taken up by intestinal epithelial cells can pass into the bloodstream. However, hydrophobic compounds and many others do not freely dissolve in the bloodstream. These compounds bind to proteins, such as albumin (Figure 36.8), that are abundant in the blood serum and by this means are carried throughout the circulatory system.

When a compound has reached the bloodstream, it is distributed to different fluids and tissues, which are often referred to as *compartments.* Some compounds are highly concentrated in their target compartments, either by binding to the target molecules themselves or by other mechanisms. Other compounds are distributed more widely (Figure 36.9). An effective drug will reach the target compartment in sufficient quantity; the concentration of the compound in the target compartment is reduced whenever the compound is distributed into other compartments.

Some target compartments are particularly hard to reach. Many compounds are excluded from the central nervous system by the *blood–brain barrier,* the tight junctions between endothelial cells that line blood vessels within the brain and spinal cord.

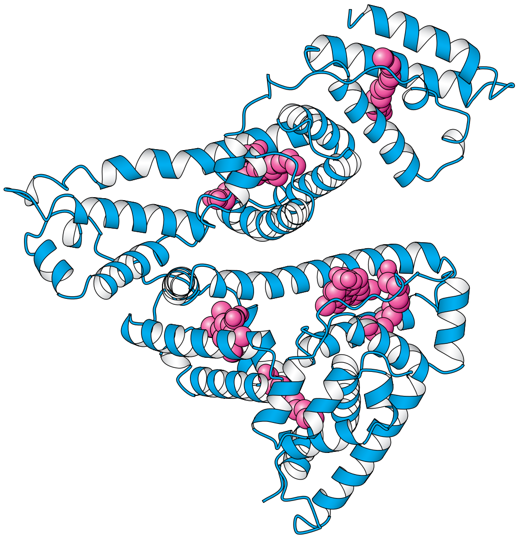


FIGURE 36.8 Structure of the drug carrier human serum albumin. Seven hydrophobic molecules (in red) are shown bound to the molecule. [Drawn from 1BKE.pdb.]

Metabolism. A final challenge to a potential drug molecule is to evade the body's defenses against foreign compounds. Many such compounds (called *xenobiotic compounds*) are released from the body in the urine or stool, often after having been metabolized to aid in excretion. This *drug metabolism* poses a considerable threat to drug effectiveness because the concentration of the desired compound decreases as it is metabolized. Thus, a rapidly metabolized compound must be administered more frequently or at higher doses.

Two of the most common pathways in xenobiotic metabolism are *oxidation* and *conjugation.* Oxidation reactions can aid excretion in at least two ways: by increasing water solubility, and thus ease of transport, and by introducing functional groups that participate in subsequent metabolic steps. These reactions are often promoted by cytochrome P450 enzymes in the liver (Section 26.4). The human genome encodes more than 50 different P450 isozymes, many of which participate in xenobiotic metabolism.

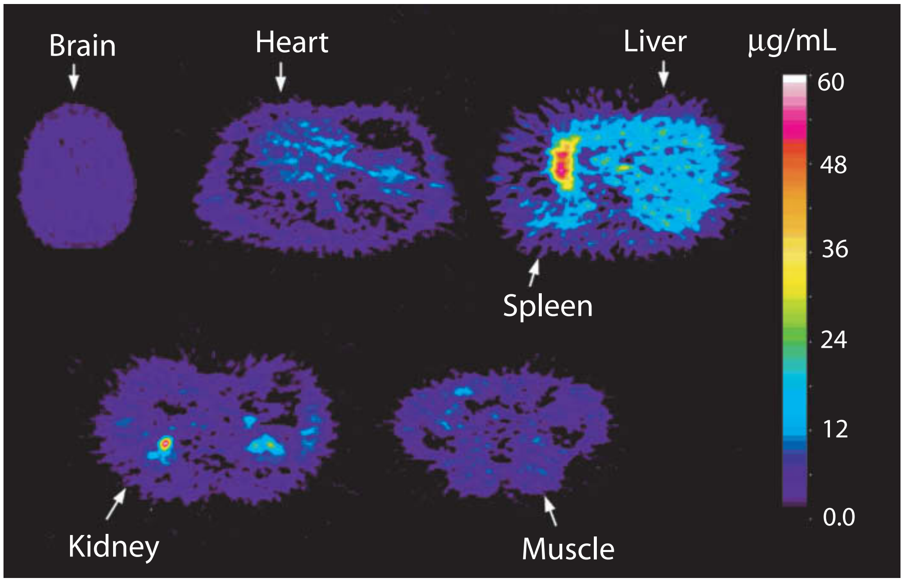
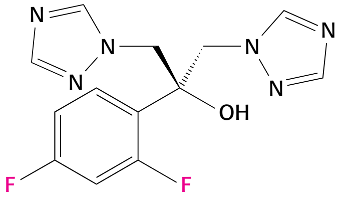
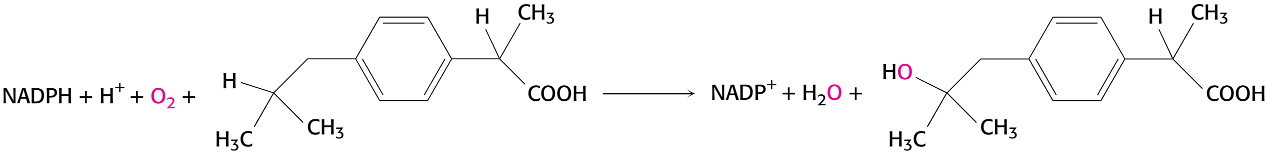


FIGURE 36.9 Distribution of the drug fluconazole. After having been taken in, compounds distribute themselves to various organs within the body. The distribution of the antifungal agent fluconazole has been monitored through the use of positron emission tomography (PET) scanning. These images were taken of a healthy human volunteer 90 minutes after injection of a dose of 5 mg kg^{-1} of fluconazole containing trace amounts of fluconazole labeled with the positron-emitting isotope ^{18}F. [Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery, Rudin, M., and Weissleder, R., "Molecular imaging in drug discovery and development," 2: 2, p. 123–131, Copyright 2003.]



Fluconazole

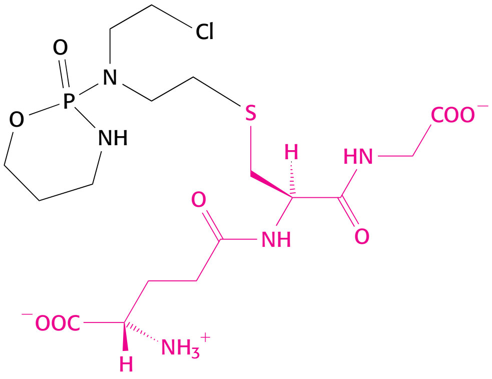
A typical reaction catalyzed by a P450 isozyme is the hydroxylation of ibuprofen (Figure 36.10).



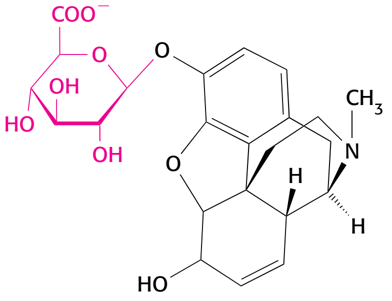
Ibuprofen

FIGURE 36.10 P450 conversion of ibuprofen. Cytochrome P450 isozymes, primarily in the liver, catalyze xenobiotic metabolic reactions such as hydroxylation. The reaction introduces an oxygen atom derived from molecular oxygen.

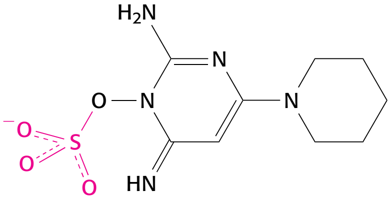
Conjugation is the addition of particular groups to the xenobiotic compound. Common groups added are glutathione (Section 20.5), glucuronic acid, and sulfate (Figure 36.11). These additions often increase water solubility and provide labels that can be recognized to target excretion. Examples of conjugation include the addition of glutathione to the anticancer drug cyclophosphamide, the addition of glucuronidate to the analgesic morphine, and the addition of a sulfate group to the hair-growth stimulator minoxidil.



Cyclophosphamide-glutathione conjugate



Morphine glucuronidate



Minoxidil sulfate

Interestingly, the sulfation of minoxidil produces a compound that is more active in stimulating hair growth than is the unmodified compound. Thus, the metabolic products of a drug, though usually less active than the drug, can sometimes be more active.

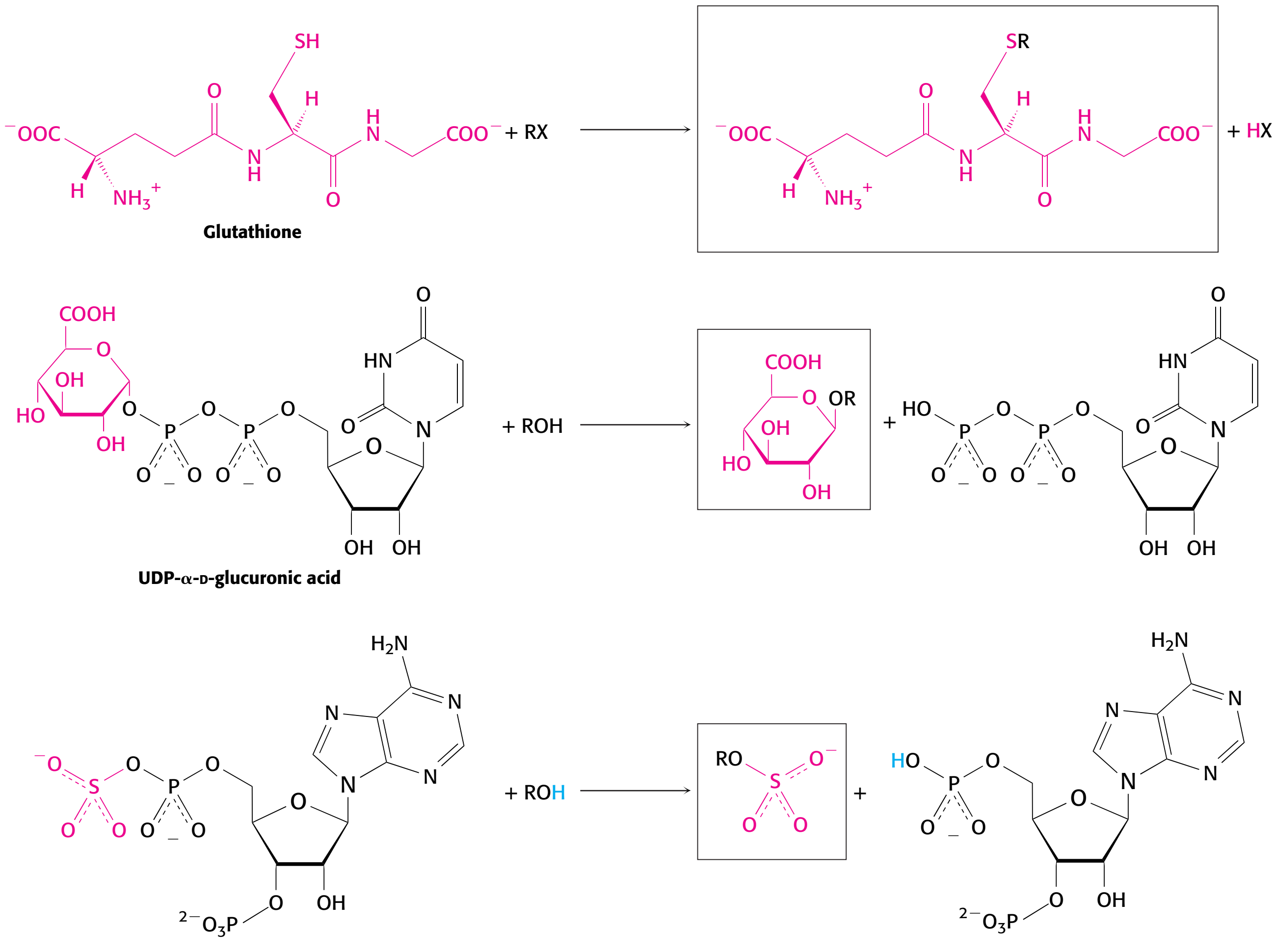


FIGURE 36.11 Conjugation reactions. Compounds that have appropriate groups are often modified by conjugation reactions. Such reactions include the addition of glutathione (top), glucuronic acid (middle), or sulfate (bottom). The conjugated product is shown boxed.

Note that an oxidation reaction often precedes conjugation because the oxidation reaction can generate modifications, such as the hydroxyl group, to which groups such as glucuronic acid can be attached. The oxidation reactions of xenobiotic compounds are often referred to as *phase I transformations,* and the conjugation reactions are referred to as *phase II transformations.* These reactions take place primarily in the liver. Because blood flows from the intestine directly to the liver through the portal vein, xenobiotic metabolism often alters drug compounds before they ever reach full circulation. This *first-pass metabolism* can substantially limit the availability of compounds taken orally.

Excretion. After compounds have entered the bloodstream, they can be removed from circulation and excreted from the body by two primary pathways. First, they can be absorbed through the kidneys and excreted in the urine. In this process, the blood passes through *glomeruli,* networks of fine capillaries in the kidney that act as filters. Compounds with molecular weights less than approximately 60,000 pass though the glomeruli. Many of the water molecules, glucose molecules, nucleotides, and other low-molecular-weight compounds that pass through the glomeruli are reabsorbed into the bloodstream, either by transporters that have broad specificities or by the passive transfer of hydrophobic molecules through membranes. Drugs and metabolites that pass through the first filtration step and are not reabsorbed are excreted.

Second, compounds can be actively transported into bile, a process that takes place in the liver. After concentration in the gall bladder, bile flows into the intestine. In the intestine, the drugs and metabolites can be excreted through the stool, reabsorbed into the bloodstream, or further degraded by

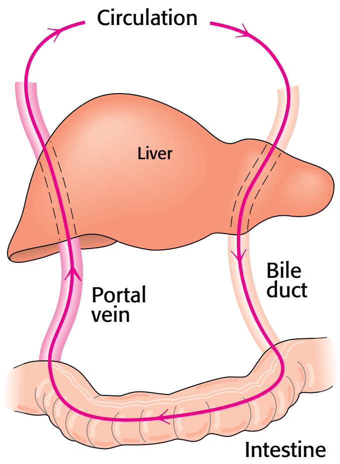


FIGURE 36.12 Enterohepatic cycling. Some drugs can move from the blood circulation to the liver, into the bile, into the intestine, to the liver, and back into circulation. This cycling decreases the rate of drug excretion.

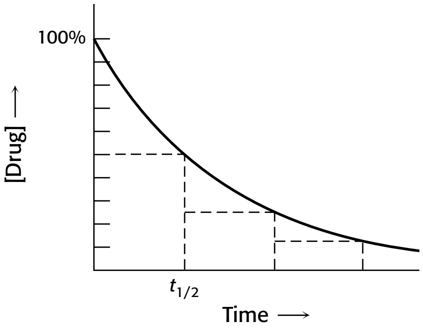


FIGURE 36.13 Half-life of drug excretion. In the case shown, the concentration of a drug in the bloodstream decreases to one-half of its value in a period of time, $t_{1/2}$, referred to as its half-life.

digestive enzymes. Sometimes, compounds are recycled from the bloodstream into the intestine and back into the bloodstream, a process referred to as *enterohepatic cycling* (Figure 36.12). This process can significantly decrease the rate of excretion of some compounds because they escape from an excretory pathway and reenter the circulation.

The kinetics of compound excretion is often complex. In some cases, a fixed percentage of the remaining compound is excreted over a given period of time (Figure 36.13). This pattern of excretion results in exponential loss of the compound from the bloodstream that can be characterized by a half-life ($t_{1/2}$). The half-life is the fixed period of time required to eliminate 50% of the remaining compound. It is a measure of how long an effective concentration of the compound remains in the system after administration. As such, the half-life is a major factor in determining how often a drug must be taken. A drug with a long half-life might need to be taken only once per day, whereas a drug with a short half-life might need to be taken three or four times per day.

Toxicity can limit drug effectiveness

An effective drug must not be so toxic that it seriously harms the person who takes it. A drug may be toxic for any of several reasons. First, it may modulate the target molecule itself *too* effectively. For example, the presence of too much of the anticoagulant drug coumadin can result in dangerous, uncontrolled bleeding and death. Second, the compound may modulate the properties of proteins that are distinct from, but related to, the target molecule itself. Compounds that are directed to one member of a family of enzymes or receptors often bind to other family members. For example, an antiviral drug directed against viral proteases may be toxic if it also inhibits proteases normally present in the body such as those that regulate blood pressure.

A compound may also be toxic if it modulates the activity of a protein unrelated to its intended target. For example, many compounds block ion channels such as the potassium channel hERG (Section 13.4), causing potentially life-threatening disturbances of cardiac rhythm. To prevent cardiac side effects, many compounds are screened for their ability to block such channels.

Finally, even if a compound is not itself toxic, its metabolic by-products may be. Phase I metabolic processes can generate damaging reactive groups in products. An important example is liver toxicity observed with large doses of the common pain reliever acetaminophen (Figure 36.14). A particular cytochrome

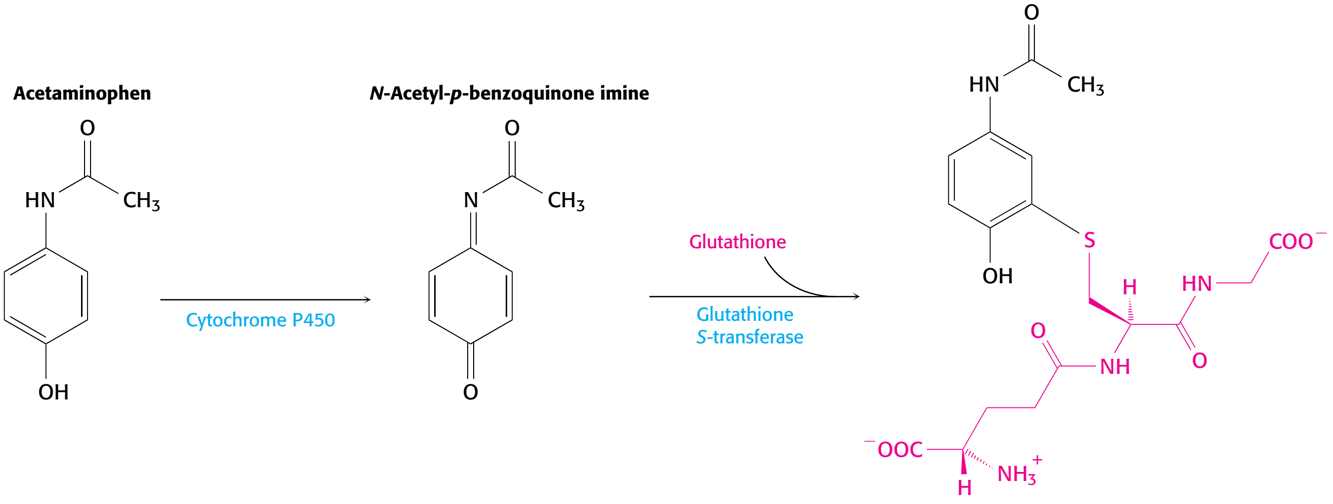


FIGURE 36.14 Acetaminophen toxicity. A minor metabolic product of acetaminophen is *N*-acetyl-*p*-benzoquinone imine. This metabolite is conjugated to glutathione. Large doses of acetaminophen can deplete liver glutathione stores.

P450 isozyme oxidizes acetaminophen to *N*-acetyl-*p*-benzoquinone imine. The resulting compound is conjugated to glutathione. With large doses, however, the liver concentration of glutathione drops dramatically, and the liver is no longer able to protect itself from this reactive compound and others. Initial symptoms of excessive acetaminophen include nausea and vomiting. Within 24 to 48 hours, symptoms of liver failure may appear. Acetaminophen poisoning accounts for about 35% of cases of severe liver failure in the United States. A liver transplant is often the only effective treatment.

The toxicity of a drug candidate can be described in terms of the *therapeutic index*. This measure of toxicity is determined through animal tests, usually with mice or rats. The therapeutic index is defined as the ratio of the dose of a compound that is required to kill one-half of the animals (referred to as the LD_{50} for "lethal dose") to a comparable measure of the effective dose, usually the EC_{50}. Thus, if the therapeutic index is 1000, then lethality is significant only when 1000 times the effective dose is administered. Analogous indices can provide measures of toxicity less severe than lethality.

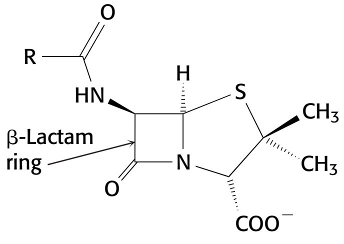
Many compounds have favorable properties in vitro, yet fail when administered to a living organism because of difficulties with ADME and toxicity. Expensive and time-consuming animal studies are required to verify that a drug candidate is not toxic, yet differences between animal species in their response can confound decisions about moving forward with a compound toward human studies. One hope is that, with more understanding of the biochemistry of these processes, scientists can develop computer-based models to replace or augment animal tests. Such models would need to accurately predict the fate of a compound inside a living organism from its molecular structure or other properties that are easily measured in the laboratory without the use of animals.

36.2 Drug Candidates Can Be Discovered by Serendipity, Screening, or Design

Traditionally, many drugs were discovered by serendipity, or chance observation. More recently, drugs have been discovered by screening collections of natural products or other compounds for molecules that have desired medicinal properties. Alternatively, scientists have designed specific drug candidates by using their knowledge about a preselected molecular target. We will examine several examples of each of these pathways to reveal common principles.

Serendipitous observations can drive drug development

Perhaps the most well-known observation in the history of drug development is Alexander Fleming's chance observation in 1928 that colonies of the bacterium *Staphylococcus aureus* died when they were adjacent to colonies of the mold *Penicillium notatum*. Spores of the mold had landed accidentally on plates growing the bacteria. Fleming soon realized that the mold produced a substance that could kill disease-causing bacteria. This discovery led to a fundamentally new approach to the treatment of bacterial infections. Howard Florey and Ernest Chain developed a powdered form of the substance, termed penicillin, that became a widely used antibiotic in the 1940s. When the structure of penicillin was elucidated in 1945, it was found to contain a four-membered β-lactam ring. This unusual feature is key to the antibacterial function of penicillin, as noted earlier (Section 8.5).



Penicillin

Three steps were crucial to fully capitalize on Fleming's discovery. First, an industrial process was developed for the production of penicillin from *Penicillium* mold on a large scale. Second, penicillin and its derivatives were chemically

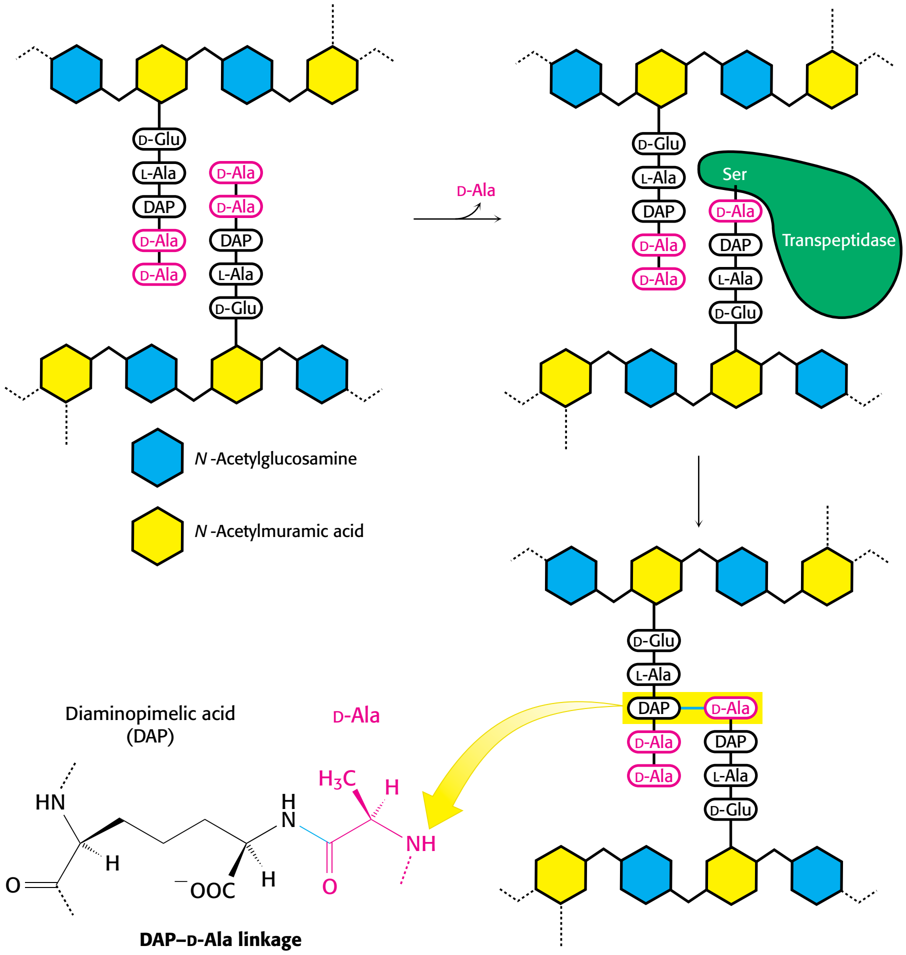
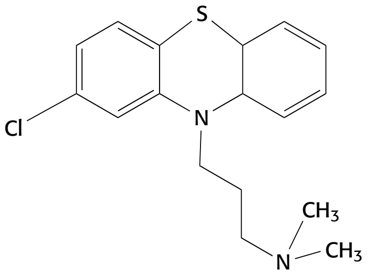


FIGURE 36.15 Mechanism of cell-wall biosynthesis disrupted by penicillin. A transpeptidase enzyme catalyzes the formation of cross-links between peptidoglycan groups. In the case shown, the transpeptidase catalyzes the linkage of D-alanine at the end of one peptide chain to the amino acid diaminopimelic acid (DAP) on another peptide chain. The diaminopimelic acid linkage (bottom left) is found in Gram-negative bacteria such as *E. coli.* Linkages of glycine-rich peptides are found in Gram-positive bacteria. Penicillin inhibits the action of the transpeptidase; so bacteria exposed to the drug have weak cell walls that are susceptible to lysis.

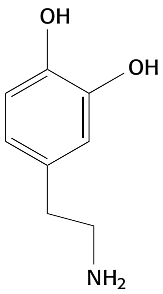
synthesized. The availability of synthetic penicillin derivatives opened the way for scientists to explore the relations between structure and function. Many such penicillin derivatives have found widespread use in medicine. Finally, in 1965, Jack Strominger and James Park independently determined that penicillin exerts its antibiotic activity by blocking a critical transpeptidase reaction in bacterial cell-wall biosynthesis (Figure 36.15), as introduced in Section 8.5.

Many other drugs have been discovered by serendipitous observations. Chlorpromazine (Thorazine), a drug used to treat psychosis, was discovered in the course of investigations directed toward the treatment of shock in surgical patients. In 1952, French surgeon Henri Laborit noticed that, after taking the compound, his patients were remarkably calm. This observation suggested that chlorpromazine could benefit psychiatric patients, and, indeed, the drug has been used for many years to treat patients with schizophrenia and bipolar disorder. The drug does have significant side effects, and its use has been largely superseded by more recently developed drugs.



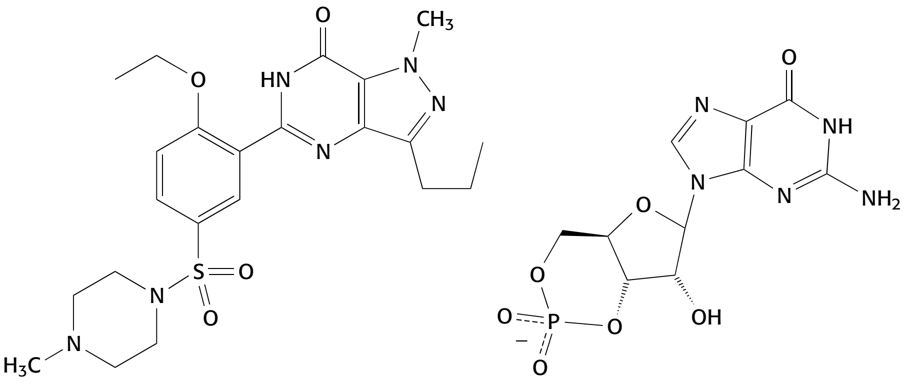
Chlorpromazine

Chlorpromazine acts by binding to receptors for the neurotransmitter dopamine and blocking them. Dopamine D2 receptors are the targets of many other psychoactive drugs. In the search for drugs with more-limited side effects, studies have been performed to correlate drug effects with biochemical parameters such as dissociation constants and binding and release rate constants.



Dopamine

A more recent example of a drug discovered by chance observation is sildenafil (Viagra). This compound was developed as an inhibitor of phosphodiesterase 5 (PDE5), an enzyme that catalyzes the hydrolysis of cGMP to GMP (Figure 36.16). The compound was intended as a treatment for hypertension and angina because cGMP plays a central role in the relaxation of



Sildenafil

cGMP

FIGURE 36.16 Sildenafil, a mimic of cGMP. Sildenafil was designed to resemble cGMP, the substrate of phosphodiesterase 5 (PDE5).

smooth-muscle cells in blood vessels (Figure 36.17). The inhibition of PDE5 was expected to increase the concentration of cGMP by blocking the pathway for its degradation. In the course of early clinical trials in Wales, some men reported unusual penile erections. Whether this chance observation by a few men was due to the compound or to other effects was unclear. However, the observation made some biochemical sense because smooth-muscle relaxation due to increased cGMP levels had been discovered to play a role in penile erection. Subsequent clinical trials directed toward the evaluation of sildenafil for erectile dysfunction were successful. This account testifies to the importance of collecting comprehensive information from clinical-trial participants. In this case, incidental observations led to a new treatment for erectile dysfunction and a multibillion-dollar-per-year drug market.

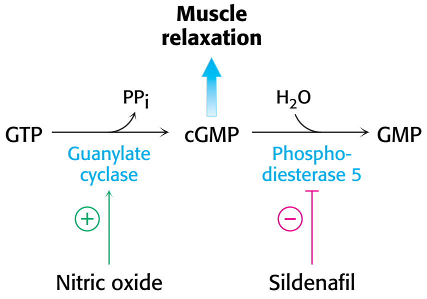
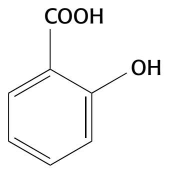


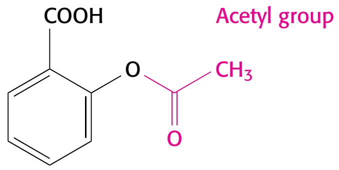
FIGURE 36.17 Muscle-relaxation pathway. Increases in nitric oxide levels stimulate guanylate cyclase, which produces cGMP. The increased cGMP concentration promotes smooth-muscle relaxation. PDE5 hydrolyzes cGMP, which lowers the cGMP concentration. The inhibition of PDE5 by sildenafil maintains elevated levels of cGMP.

Natural products are a valuable source of drugs and drug leads

No drug is as widely used as aspirin. Observers at least as far back as Hippocrates (∼400 B.C.) have noted the use of extracts from the bark and leaves of the willow tree for pain relief. In 1829, a mixture called *salicin* was isolated from willow bark. Subsequent analysis identified salicylic acid as the active component of this mixture. Salicylic acid was formerly used to treat pain, but this compound often irritated the stomach. Several investigators attempted to find a means to neutralize salicylic acid. Felix Hoffmann, a chemist working at the German company Bayer, developed a less-irritating derivative by treating salicylic acid with a base and acetyl chloride. This derivative, acetylsalicylic acid, was named *aspirin* from "a" for acetyl chloride, "spir" for *Spiraea ulmaria* (meadowsweet, a flowering plant that also contains salicylic acid), and "in" (a common ending for drugs). Each year, approximately 35,000 tons of aspirin are taken worldwide, nearly the weight of the *Titanic.*



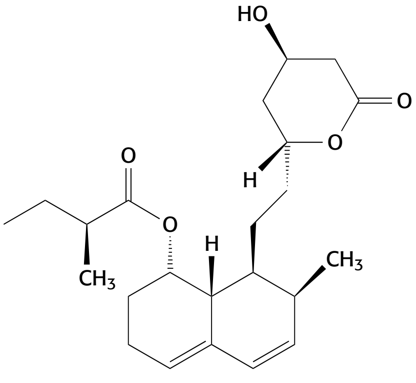
Salicylic acid



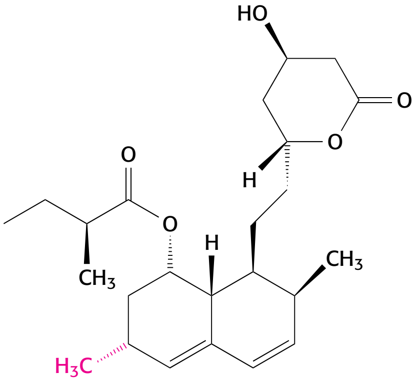
Aspirin (acetylsalicylic acid)

As discussed in Chapter 12, the acetyl group in aspirin is transferred to the side chain of a serine residue that lies along the path to the active site of the cyclooxygenase component of prostaglandin H_2 synthase (Figure 12.24). In this position, the acetyl group blocks access to the active site. Thus, even though aspirin binds in the same pocket on the enzyme as salicylic acid, the acetyl group of aspirin dramatically increases its effectiveness as a drug. This account illustrates the value of screening extracts from plants and other materials that are believed to have medicinal properties for active compounds. The large number of herbal and folk medicines are a treasure trove of new drug leads.

Let us consider another example, which has also had a significant impact on the clinical management of individuals with cardiovascular disease. More than 100 years ago, a fatty, yellowish material was discovered on the arterial walls of patients who had died of vascular disease. The presence of the material was termed *atheroma* from the Greek word for porridge. This material proved to be cholesterol. The Framingham heart study, initiated in 1948, documented



Compactin



Lovastatin

a correlation between high blood-cholesterol levels and high mortality rates from heart disease. This observation led to the notion that blocking cholesterol synthesis might lower blood-cholesterol levels and, in turn, lower the risk of heart disease. Initial attempts at blocking cholesterol synthesis focused on steps near the end of the pathway. However, these efforts were abandoned because the accumulation of the insoluble substrate for the inhibited enzyme led to the development of cataracts and other side effects. Investigators eventually identified a more-favorable target—namely, the enzyme HMG-CoA reductase (Section 26.2). This enzyme acts on a substrate, HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A), that does not accumulate, because it is water-soluble and can be utilized by other pathways.

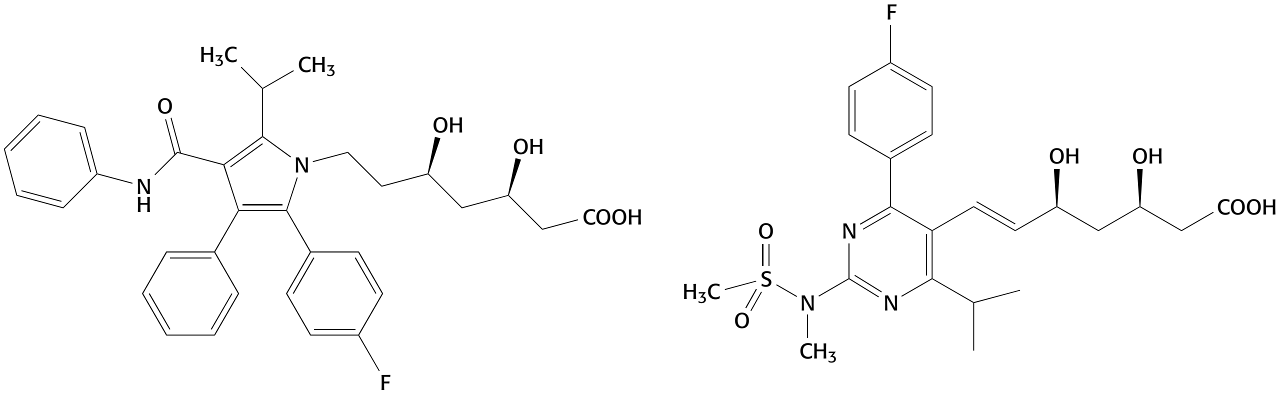
A promising natural product, compactin, was discovered in a screen of compounds from a fermentation broth from *Penicillium citrinum* in a search for antibacterial agents. In some, but not all, animal studies, compactin was found to inhibit HMG-CoA reductase and to lower serum cholesterol levels. In 1982, a new HMG-CoA reductase inhibitor was discovered in a fermentation broth from *Aspergillus cereus*. This compound, now called lovastatin, was found to be structurally very similar to compactin, bearing one additional methyl group.

In clinical trials, lovastatin significantly reduced serum cholesterol levels with few side effects. Most side effects could be prevented by treatment with mevalonate (the product of HMG-CoA reductase), indicating that the side effects were likely due to the highly effective blocking of HMG-CoA reductase. One notable side effect is muscle pain or weakness (termed *myopathy*), although its cause remains to be fully established. After many studies, the Food and Drug Administration (FDA) approved lovastatin for treating high serum cholesterol levels.

A structurally related HMG-CoA reductase inhibitor was later shown to cause a statistically significant decrease in deaths due to coronary heart disease. This result validated the benefits of lowering serum cholesterol levels. Further mechanistic analysis revealed that the HMG-CoA reductase inhibitor acts not only by lowering the rate of cholesterol biosynthesis, but also by inducing the expression of the low-density-lipoprotein (LDL) receptor (Section 26.3). Cells with such receptors remove LDL particles from the bloodstream, and so these particles cannot contribute to atheroma.

Screening libraries of synthetic compounds expands the opportunity for identification of drug leads

Lovastatin and its relatives are either natural products or readily derived from natural products. After the discovery of these compounds, totally synthetic molecules were developed that are more-potent inhibitors of HMG-CoA reductase (Figure 36.18). These compounds were found to be effective



Atorvastatin **Rosuvastatin**

FIGURE 36.18 Synthetic statins. Atorvastatin (Lipitor) and rosuvastatin (Crestor) are completely synthetic drugs that inhibit HMG-CoA reductase.

at lower dose levels, reducing side effects. The original HMG-CoA reductase inhibitors or their precursors were found by screening libraries of natural products. More recently, drug developers have tried screening large libraries of both natural products and purely synthetic compounds prepared in the course of many drug-development programs. Under favorable circumstances, hundreds of thousands or even millions of compounds can be tested in this process, termed *high-throughput screening.* Compounds in these libraries can be synthesized one at a time for testing. An alternative approach is to synthesize a large number of structurally related compounds that differ from one another at only one or a few positions all at once. This approach is often termed *combinatorial chemistry.* Here, compounds are synthesized with the use of the same chemical reactions but a variable set of reactants. Suppose that a molecular scaffold is constructed with two reactive sites and that 20 reactants can be used in the first site and 40 reactants can be used in the second site. A total of 20 × 40 = 800 possible compounds can be produced.

A key method in combinatorial chemistry is *split-pool synthesis* (Figure 36.19). The method depends on solid-phase synthetic methods, first developed for the synthesis of peptides (Section 3.5). Compounds are synthesized on small beads. Beads containing an appropriate starting *scaffold* are produced and divided (split) into *n* sets, with *n* corresponding to the number of building blocks to be used at one site. Reactions adding the reactants at the

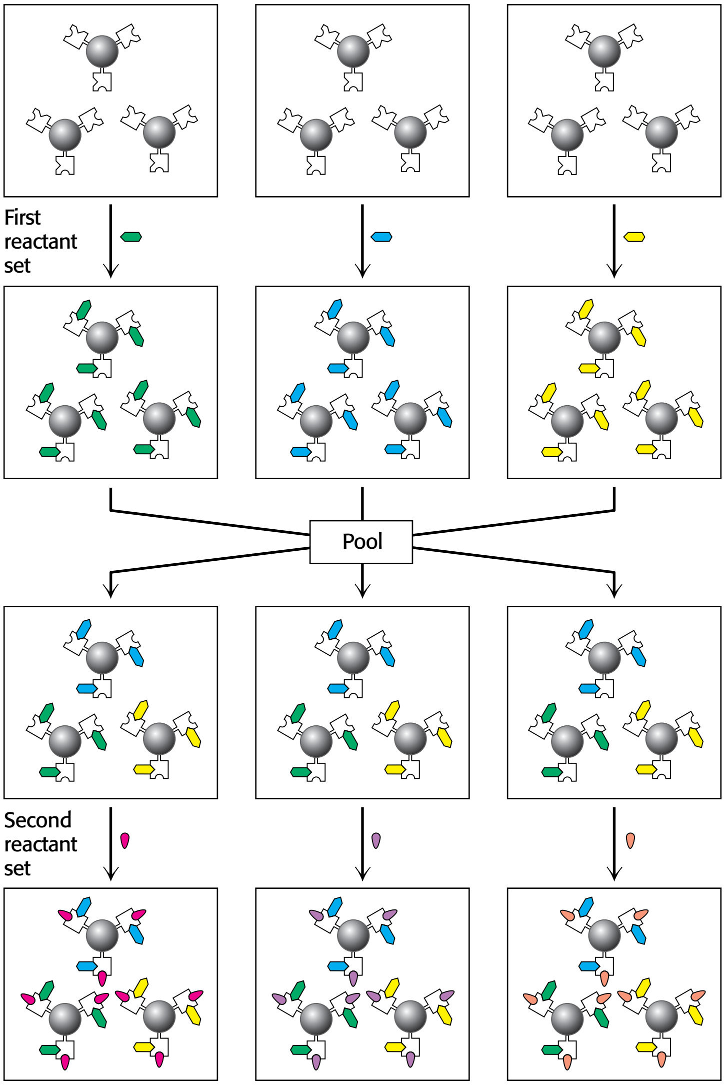


FIGURE 36.19 Split-pool synthesis. Reactions are performed on beads. Each of the reactions with the first set of reactants is performed on a separate set of beads. The beads are then pooled, mixed, and split into sets. The second set of reactants is then added. Many different compounds will be produced, but all of the compounds on a single bead will be identical.

first site are run, and the beads are isolated by filtration. The n sets of beads are then combined (pooled), mixed, and split again into m sets, with m corresponding to the number of reactants to be used at the second site. Reactions adding these m reactants are run, and the beads are again isolated. The important result is that each bead contains only one compound, even though the entire library of beads contains many. Furthermore, although only $n + m$ reactions were run, $n \times m$ compounds are produced. With the preceding values for n and m, $20 + 40 = 60$ reactions produce $20 \times 40 = 800$ compounds. In some cases, assays can be performed directly with the compounds still attached to the bead to find compounds with desired properties (Figure 36.20). Alternatively, each bead can be isolated and the compound can be cleaved from the bead to produce free compounds for analysis. After an interesting compound has been identified, analytical methods of various types must be used to identify which of the $n \times m$ compounds is present.

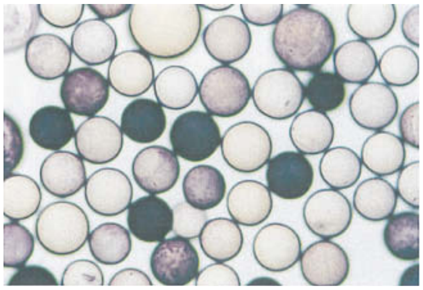


FIGURE 36.20 Screening a library of synthesized carbohydrates. A small combinatorial library of carbohydrates synthesized on the surface of 130-mm beads is screened for carbohydrates that are bound tightly by a lectin from peanuts. Beads that have such carbohydrates are darkly stained through the action of an enzyme linked to the lectin. [From Figure 3 in Liang et al., "Polyvalent binding to carbohydrates immobilized on an insoluble resin," Proceedings of the National Academy of Sciences, USA, vol. 94, pp. 10554–10559, September 1997. Copyright 2004 National Academy of Sciences, USA.]

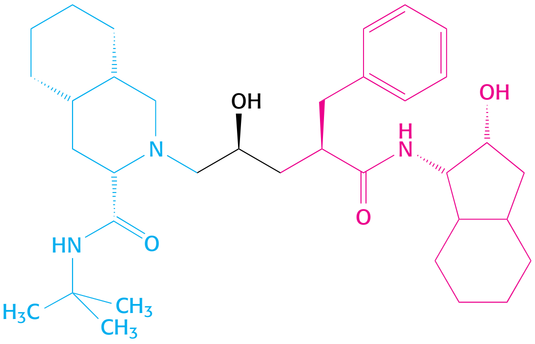
Note that the "universe" of druglike compounds is vast. More than an estimated 10^{40} compounds are possible with molecular weights less than 750. Thus, even with "large" libraries of millions of compounds, only a tiny fraction of the chemical possibilities are present for study.

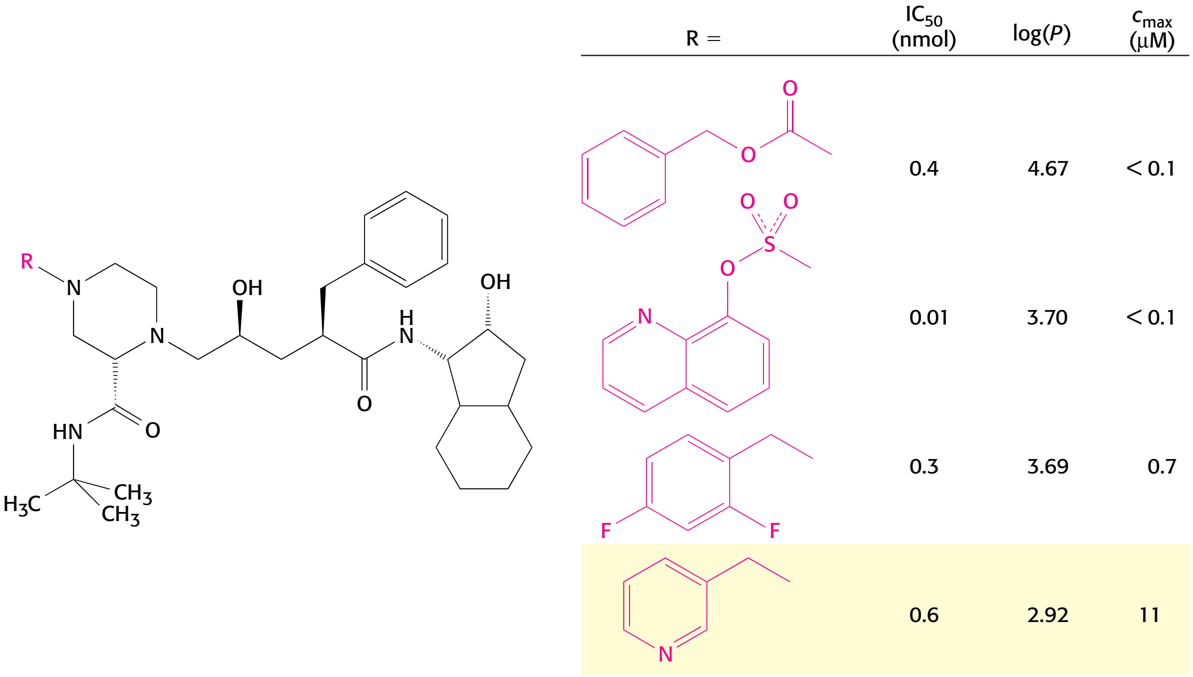
Drugs can be designed on the basis of three-dimensional structural information about their targets

Many drugs bind to their targets in a manner reminiscent of Emil Fischer's lock and key (Figure 8.8). Therefore, we should be able to design a key, given enough knowledge about the shape and chemical composition of the lock. In the idealized case, we would like to design a small molecule that is complementary in shape and electronic structure to a target protein so that it binds effectively to the targeted site. Despite our ability to determine three-dimensional structures rapidly, the achievement of this goal remains in the future. Designing stable compounds *from scratch* that have the correct shape and other properties to fit precisely into a binding site is difficult because predicting the structure that will best fit into a binding site is difficult. Prediction of binding affinity requires a detailed understanding of the interactions between a compound and its binding partner *and* of the interactions between the compound and the solvent when the compound is free in solution.

Nonetheless, *structure-based drug design* has proved to be a powerful tool in drug development. Among its most prominent successes has been the development of drugs that inhibit the protease from the human immunodeficiency virus (HIV; Section 34.4). Consider the development of the protease inhibitor indinavir (Crixivan; Figure 9.19). Two sets of promising inhibitors that had high potency but poor solubility and bioavailability were discovered. X-ray crystallographic analysis and molecular-modeling findings suggested that a hybrid of these two inhibitor sets might have both high potency and improved bioavailability (Figure 36.21). The synthesized hybrid compound did show improvements but required further optimization. The structural data suggested one point where modifications could be tolerated. A series of compounds were prepared and tested for their ability to inhibit the protease (Figure 36.22). These data are a demonstration of a *structure–activity relationship (SAR)*; they provide an opportunity to correlate structure with function and guide design of further molecules. The most-active compound showed poor bioavailability, but one of the other compounds (highlighted in yellow in Figure 36.22) showed good bioavailability and acceptable activity. The maximum serum concentration available through oral administration was significantly higher than the levels required to suppress

FIGURE 36.21 Initial design of an HIV protease inhibitor. This compound was designed by combining part of one compound with good inhibition activity but poor solubility (shown in red) with part of another compound with better solubility (shown in blue).





R =	IC_{50} (nmol)	log(*P*)	c_{max} (μM)
	0.4	4.67	< 0.1
	0.01	3.70	< 0.1
	0.3	3.69	0.7
	0.6	2.92	11

FIGURE 36.22 Compound optimization. Four compounds are evaluated for characteristics including the IC_{50}, log(*P*), and c_{max} (the maximal concentration of compound present) measured in the serum of dogs. The compound shown at the bottom (highlighted in yellow) has the weakest inhibitory power (measured by IC_{50}) but by far the best bioavailability (measured by c_{max}). This compound was selected for further development, leading to the drug indinavir (Crixivan).

replication of the virus. This drug, as well as other protease inhibitors developed at about the same time, has been used in combination with other drugs to treat AIDS with much more encouraging results than had been obtained previously (Figure 36.23).

Aspirin targets the cyclooxygenase site in prostaglandin H_2 synthase, as discussed earlier. Animal studies suggested that mammals contain not one but two distinct cyclooxygenase enzymes, both of which are targeted by aspirin. The more recently discovered enzyme, cyclooxygenase 2 (COX2), is expressed primarily as part of the inflammatory response, whereas cyclooxygenase 1 (COX1) is expressed more generally. These observations suggested that a cyclooxygenase inhibitor that was specific for COX2 might be able to reduce inflammation in conditions such as arthritis without producing the gastric and other side effects associated with aspirin.

The amino acid sequences of COX1 and COX2 were deduced from cDNA cloning studies. These sequences are more than 60% identical, clearly indicating that the enzymes have the same overall structure. Nevertheless, there are some differences in the residues around the aspirin-binding site. X-ray crystallography revealed that an extension of the binding pocket was present in COX2, but absent in COX1. This structural difference suggested

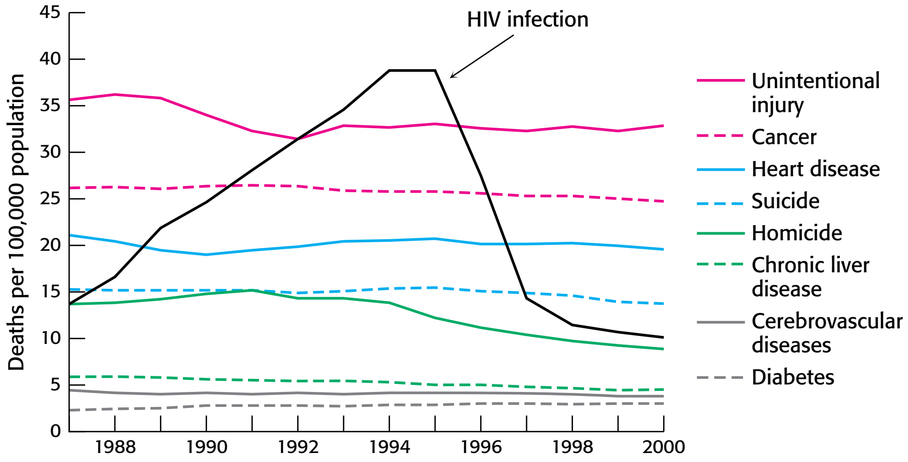


FIGURE 36.23 The effect of anti-HIV drug development. Death rates from HIV infection (AIDS) reveal the tremendous effect of HIV protease inhibitors and their use in combination with inhibitors of HIV reverse transcriptase. The death rates in this graph are from the leading causes of death among persons 24 to 44 years old in the United States. [Data from Centers for Disease Control.]

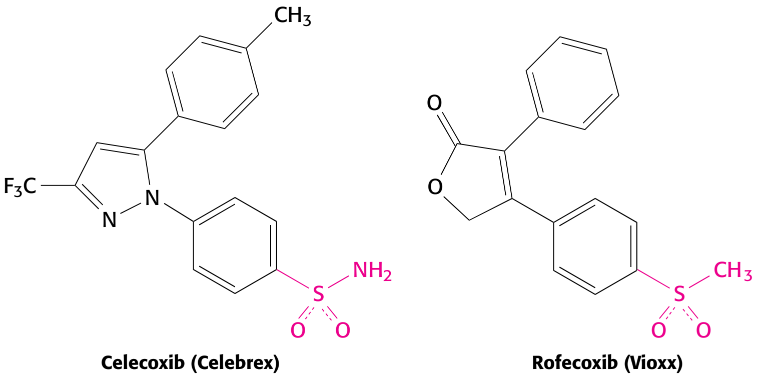


FIGURE 36.24 COX2-specific inhibitors. These compounds have protuberances (shown in red) that fit into a pocket in the COX2 isozyme but sterically clash with the COX1 isozyme.

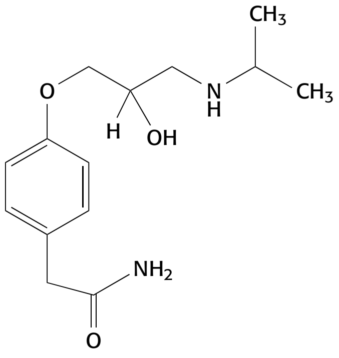
a strategy for constructing COX2-specific inhibitors—namely, to synthesize compounds that had a protuberance that would fit into the pocket in the COX2 enzyme. Such compounds were designed and synthesized and then further refined to produce effective drugs familiar as Celebrex and Vioxx (Figure 36.24). Vioxx was subsequently withdrawn from the market because some people experienced adverse effects. Some of these effects may be due to the inhibition of COX2, the intended target. Thus, although the development of these drugs is a triumph for structure-based drug design, these outcomes highlight the fact that the inhibition of important enzymes can lead to complex physiological responses.

36.3 Analyses of Genomes Hold Great Promise for Drug Discovery

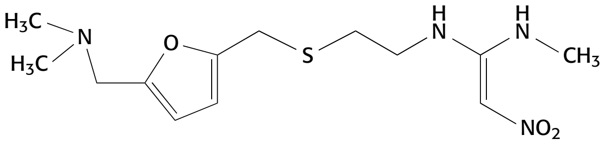
The completion of the sequencing of the human and other genomes is a potentially powerful driving force for the development of new drugs. Genomic sequencing and analysis projects have vastly increased our knowledge of the proteins encoded by the human genome. This new source of knowledge may greatly accelerate early stages of the drug-development process or even allow drugs to be tailored to the individual patient.

Potential targets can be identified in the human proteome

The human genome encodes approximately 21,000 proteins, not counting the variation produced by alternative mRNA splicing and posttranslational modifications. Many of these proteins are potential drug targets, in particular those that are enzymes or receptors and have significant biological effects when activated or inhibited. Several large protein families are particularly rich sources of targets. For example, the human genome includes genes for more than 500 protein kinases that can be recognized by comparing the deduced amino acid sequences. Many of these kinases are known to play a role in the progression of a variety of diseases. For example, Bcr-Abl kinase, a dysregulated kinase formed from a specific chromosomal defect, is known to contribute to certain leukemias and is the target of the drug imatinib mesylate (Gleevec; Section 14.5). Some of the other protein kinases undoubtedly play central roles in particular cancers as well. Similarly, the human genome encodes approximately 800 7TM receptors (Section 14.1), of which approximately 350 are odorant receptors. Many of the remaining 7TM receptors are known or potential drug targets. For example, the beta-blockers, which are widely used to treat hypertension, target the β_1-adrenergic receptor, and the antiulcer medication ranitidine (Zantac) targets the histamine H_2 receptor (a 7TM receptor that participates in the control of gastric acid secretion).



Atenolol

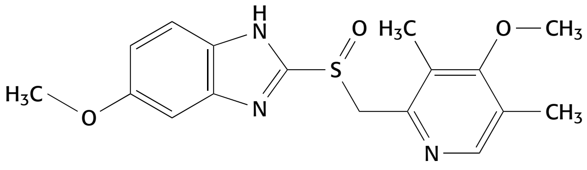


Ranitidine

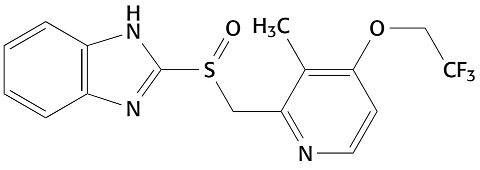
Novel proteins that are not part of large families already supplying drug targets can be more readily identified through the use of genomic information. There are a number of ways to identify proteins that could serve as targets of drug-development programs. One way is to look for changes in expression patterns, protein localization, or posttranslational modifications in cells from disease-afflicted organisms. Another is to perform studies of tissues or cell types in which particular genes are expressed. Such analyses of the human genome should increase the number of actively pursued drug targets.

Animal models can be developed to test the validity of potential drug targets

The genomes of a number of model organisms have now been sequenced. The most important of these genomes for drug development is that of the mouse. Remarkably, the mouse and human genomes are approximately 85% identical in sequence, and more than 98% of all human genes have recognizable mouse counterparts. Mouse studies provide drug developers with a powerful tool—the ability to disrupt (“knock out”) specific genes in the mouse (Section 5.4). If disruption of a gene has a desirable effect, then the product of this gene may represent a promising drug target. The utility of this approach has been demonstrated retrospectively. For example, disruption of the gene for the α subunit of the H^+–K^+ ATPase, the key protein for secreting acid into the stomach, produces mice with a gastric pH of 6.9. Under similar conditions, their wild-type counterparts produce a gastric pH of 3.2. This protein is the target of the drugs omeprazole (Prilosec) and lansoprazole (Prevacid and Takepron), used for treating gastroesophageal reflux disease.



Omeprazole



Lansoprazole

Several large-scale efforts are underway to generate hundreds or thousands of mouse strains, each having a different gene disrupted. The phenotypes of these mice are a good indication of whether the protein encoded by a disrupted gene is a promising drug target. This approach allows drug developers to evaluate potential targets without any preconceived notions regarding physiological function.

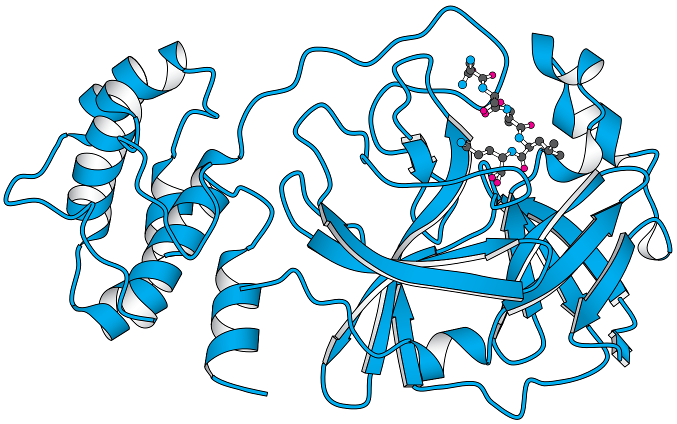


FIGURE 36.25 Emerging drug target. The structure of a protease from the coronavirus that causes SARS (severe acute respiratory syndrome) is shown bound to an inhibitor. This structure was determined less than a year after the identification of the virus. [Drawn from 1P9S.pdb.]

Potential targets can be identified in the genomes of pathogens

Human proteins are not the only important drug targets. Drugs such as penicillin and HIV protease inhibitors act by targeting proteins within a pathogen. The genomes of hundreds of pathogens have now been sequenced, and these genome sequences can be mined for potential targets.

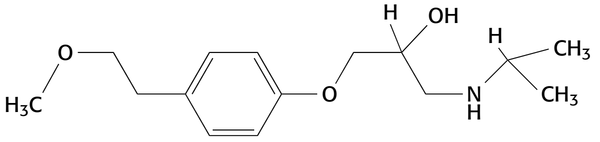
New antibiotics are needed to combat bacteria that are resistant to many existing antibiotics. One approach seeks proteins essential for cell survival that are conserved in a wide range of bacteria. Drugs that inactivate such proteins are expected to be broad-spectrum antibiotics, useful for treating infections from any of a range of different bacteria. One such protein is peptide deformylase, the enzyme that removes formyl groups present at the amino termini of bacterial proteins immediately after translation (Figure 30.19).

Alternatively, a drug may be needed against a specific pathogen. An example of such a pathogen is the strain of coronavirus responsible for severe acute respiratory syndrome (SARS). Within one month of the recognition of this emerging disease, investigators had isolated the virus that causes the syndrome, and, within weeks, completely sequenced its 29,751-base genome. This sequence revealed the presence of a gene encoding a viral protease, known to be essential for viral replication from studies of other members of the coronavirus family to which the SARS virus belongs. This structure has opened the possibility for specific antiviral treatments for this, and related, viruses (Figure 36.25).

Genetic differences influence individual responses to drugs

Many drugs are not effective in everyone, often because of genetic differences among individuals. Nonresponding persons may have slight differences in either a drug's target molecule or the proteins taking part in drug transport and metabolism. The goal of the emerging fields of pharmacogenetics and pharmacogenomics is to design drugs that either act more consistently from person to person or are tailored to individual persons with particular genotypes.

Drugs such as metoprolol that target the β_1-adrenergic receptor are popular treatments for hypertension. These drugs are often referred to as "beta-blockers."



Metoprolol

However, some individuals do not respond well to beta-blockers. Two variants of the gene coding for the β_1-adrenergic receptor are common in the U.S. population. The most common allele has serine in position 49 and arginine in position 389. In some persons, however, glycine replaces one or the other of these residues. In clinical studies, participants with two copies of the most common allele responded well to metoprolol: their daytime diastolic blood pressure was reduced by 14.7 ± 2.9 mm Hg on average. In contrast, participants with one variant allele showed a smaller reduction in blood pressure, and the drug had no significant effect on participants with two variant alleles (Figure 36.26). These observations suggest the potential utility of genotyping

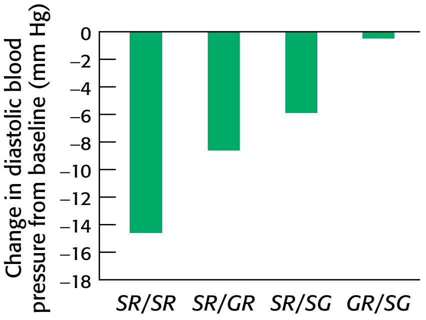
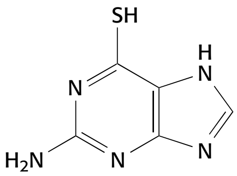


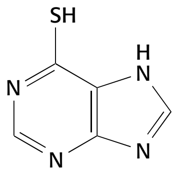
FIGURE 36.26 Phenotype–genotype correlation. Average changes in diastolic blood pressure on treatment with metoprolol. Persons with two copies of the most common allele ($S_{49}R_{389}$) showed significant decreases in blood pressure. Those with one variant allele (*GR* or *SG*) showed more modest decreases, and those with two variant alleles (*GR/SG*) showed no decrease. [Data from J. A. Johnson et al., *Clin. Pharmacol. Ther.* 74:44–52, 2003.]

individual patients at these positions. One could then predict whether treatment with metoprolol or other beta-blockers is likely to be effective.

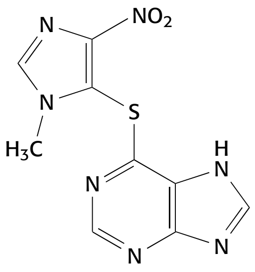
Given the importance of ADME and toxicity properties in determining drug efficacy, it is not surprising that variations in proteins participating in drug transport and metabolism can alter a drug's effectiveness. An important example is the use of thiopurine drugs such as 6-thioguanine, 6-mercaptopurine, and azothioprine to treat diseases including leukemia, immune disorders, and inflammatory bowel disease.



6-Thioguanine

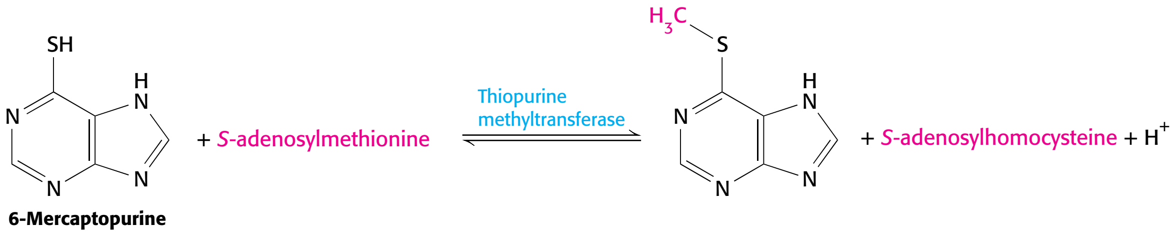


6-Mercaptopurine



Azathioprine

A minority of patients who are treated with these drugs show signs of toxicity at doses that are well tolerated by most patients. These differences between patients are due to rare variations in the gene encoding the xenobiotic-metabolizing enzyme thiopurine methyltransferase, which adds a methyl group to sulfur atoms.



The variant enzyme is less stable. Patients with these variant enzymes can build up toxic levels of the drugs if appropriate care is not taken. Thus, genetic variability in an enzyme participating in drug metabolism plays a large role in determining the variation in the tolerance of different persons to particular drug levels. Many other drug-metabolism enzymes and drug-transport proteins have been implicated in controlling individual reactions to specific drugs. The identification of the genetic factors will allow a deeper understanding of why some drugs work well in some individuals but poorly in others. In the future, doctors may examine a patient's genes to help plan drug-therapy programs.

36.4 The Clinical Development of Drugs Proceeds Through Several Phases

In the United States, the FDA requires demonstration that drug candidates be effective and safe before they may be used in human beings on a large scale. This requirement is particularly true for drug candidates that are to be taken by people who are relatively healthy. More side effects are acceptable for drug candidates intended to treat significantly ill patients such as those with serious forms of cancer, where there are clear, unfavorable consequences for not having an effective treatment.

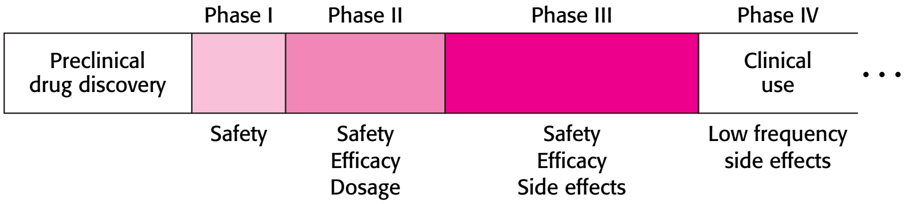

Phase I
Phase II
Phase III
Phase IV
Preclinical drug discovery
Clinical use
Safety
Safety Efficacy Dosage
Safety Efficacy Side effects
Low frequency side effects

FIGURE 36.27 Clinical-trial phases. Clinical trials proceed in phases examining safety and efficacy in increasingly large groups.

Clinical trials are time consuming and expensive

Clinical trials test the effectiveness and potential side effects of a candidate drug before it is approved by the FDA for general use. These trials proceed in at least three phases (Figure 36.27). In Phase I, a small number (typically from 10 to 100) of usually healthy volunteers take the drug for an initial study of safety. These volunteers are given a range of doses and are monitored for signs of toxicity. The efficacy of the drug candidate is not specifically evaluated.

In Phase II, the efficacy of the drug candidate is tested in a small number of persons who might benefit from the drug. Further data regarding the drug's safety are obtained. Such trials are often controlled and double-blinded. In a controlled study, subjects are divided randomly into two groups. Subjects in the treatment group are given the treatment under investigation. Subjects in the control group are given either a *placebo*—that is, a treatment such as sugar pills known to not have intrinsic value—or the best standard treatment available, if withholding treatment altogether would be unethical. In a double-blinded study, neither the subjects nor the researchers know which subjects are in the treatment group and which are in the control group. A double-blinded study prevents bias in the course of the trial. When the trial has been completed, the assignments of the subjects into treatment and control groups are unsealed and the results for the two groups are compared. A variety of doses are often investigated in Phase II trials to determine which doses appear to be free of serious side effects and which doses appear to be effective.

The power of the *placebo effect*—that is, the tendency to perceive improvement in a subject who believes that he or she is receiving a potentially beneficial treatment—should not be underestimated. In a study of arthroscopic surgical treatment for knee pain, for example, subjects who were led to believe that they had received surgery through the use of videotapes and other means showed the same level of improvement, on average, as subjects who actually received the procedure.

In Phase III, similar studies are performed on a larger and more diverse population. This phase is intended to more firmly establish the efficacy of the drug candidate and to detect side effects that may develop in a small percentage of the subjects who receive treatment. Thousands of subjects may participate in a typical Phase III study.

Clinical trials can be extremely costly. Hundreds or thousands of patients must be recruited and monitored for the duration of the trial. Many physicians, nurses, clinical pharmacologists, statisticians, and others participate in the design and execution of the trial. Costs can run from tens of millions to hundreds of millions of dollars. Extensive records must be kept, including documentation of any adverse reactions. These data are compiled and submitted to the FDA. The full cost of developing a drug is currently estimated to be more than $800 million.

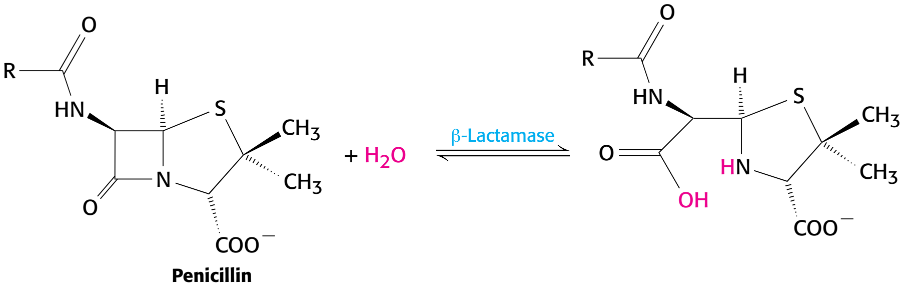
Even after a drug has been approved and is in use, difficulties can arise. Clinical trials run after a drug has entered the market, referred to as Phase IV studies, are designed to identify low-frequency side effects that may only emerge after widespread or long-term use. As mentioned earlier, rofecoxib (Vioxx), for example, was withdrawn from the market after significant cardiac side effects were detected in Phase IV clinical trials. Such events highlight the necessity for users of any drug to balance beneficial effects against potential risks.

The evolution of drug resistance can limit the utility of drugs for infectious agents and cancer

Many drugs are used for long periods of time without any loss of effectiveness. However, in some cases, particularly for the treatment of cancer or infectious diseases, drug treatments that were initially effective become less so. In other words, the disease becomes resistant to the drug therapy. Why does this resistance develop? Infectious diseases and cancer have a common feature—namely, that an affected person contains many cells (or viruses) that can mutate and reproduce. These conditions are necessary for evolution to take place. Thus, an individual microorganism or cancer cell may, by chance, have a genetic variation that makes it more suitable for growth and reproduction in the presence of the drug. These microorganisms or cells are more fit than others in their population, and they will tend to take over the population. As the selective pressure due to the drug is continually applied, the population of microorganisms or cancer cells will tend to become more and more resistant to the presence of the drug. Note that resistance can develop by a number of mechanisms.

The HIV protease inhibitors discussed earlier provide an important example of the evolution of drug resistance. Retroviruses are very well suited to this sort of evolution because reverse transcriptase carries out replication without a proofreading mechanism. In a genome of approximately 9750 bases, each possible single point mutation is estimated to appear in a virus particle more than 1000 times per day in each infected person. Many multiple mutations also occur. Most of these mutations either have no effect or are detrimental to the virus. However, a few of the mutant virus particles encode proteases that are less susceptible to inhibition by the drug. In the presence of an HIV protease inhibitor, these virus particles will tend to replicate more effectively than does the population at large. With the passage of time, the less-susceptible viruses will come to dominate the population and the virus population will become resistant to the drug.

Pathogens may become resistant to antibiotics by completely different mechanisms. Some pathogens contain enzymes that inactivate or degrade specific antibiotics. For example, many bacteria are resistant to β-lactams such as penicillin because they contain β-lactamase enzymes. These enzymes hydrolyze the β-lactam ring and render the drugs inactive.



Many of these enzymes are encoded in plasmids, small circular pieces of DNA often carried by bacteria. Many plasmids are readily transferred from one bacterial cell to another, transmitting the capability for antibiotic resistance. Plasmid transfer thus contributes to the spread of antibiotic resistance, a major health-care challenge. On the other hand, plasmids have been harnessed for use in recombinant DNA methods (Section 5.2).

Drug resistance commonly emerges in the course of cancer treatment. Cancer cells are characterized by their ability to grow rapidly without the constraints that apply to normal cells. Many drugs used for cancer chemotherapy inhibit processes that are necessary for this rapid cell growth.

However, individual cancer cells may accumulate genetic changes that mitigate the effects of such drugs. These altered cancer cells will tend to grow more rapidly than others and will become dominant within the cancer-cell population. This ability of cancer cells to mutate quickly has been a challenge to one of the major breakthroughs in cancer treatment: the development of inhibitors for proteins specific to cancer cells present in certain leukemias (Section 14.5). For example, tumors become undetectable in patients treated with imatinib mesylate, which is directed against the Bcr-Abl protein kinase. Unfortunately, the tumors of many of the patients treated with imatinib mesylate recur after a period of years. In many of these cases, mutations have altered the Bcr-Abl protein so that it is no longer inhibited by the concentrations of imatinib mesylate used in therapy.

Cancer patients often take multiple drugs concurrently in the course of chemotherapy and, in many cases, cancer cells become simultaneously resistant to many or all of them. This multiple-drug resistance can be due to the proliferation of cancer cells that overexpress a number of ABC transporter proteins that pump drugs out of the cell (Section 13.2). Thus, cancer cells can evolve drug resistance by over expressing normal human proteins or by modifying proteins responsible for the cancer phenotype.

SUMMARY

36.1 The Development of Drugs Presents Huge Challenges

Most drugs act by binding to enzymes or receptors and modulating their activities. To be effective, drugs must bind to these targets with high affinity and specificity. However, compounds with the desired affinity and specificity do not necessarily make suitable drugs. Most compounds are poorly absorbed or rapidly excreted from the body or they are modified by metabolic pathways that target foreign compounds. Consequently, when taken orally, these compounds do not reach their targets at appropriate concentrations for a sufficient period of time. A drug's properties related to its absorption, distribution, metabolism, and excretion are called ADME properties. Oral bioavailability is a measure of a drug's ability to be absorbed; it is the ratio of the peak concentration of a compound given orally to the peak concentration of the same dose directly injected. The structure of a compound can affect its bioavailability in complicated ways, but generalizations called Lipinski's rules provide useful guidelines. Drug-metabolism pathways include oxidation by cytochrome P450 enzymes (phase I metabolism) and conjugation to glutathione, glucuronic acid, and sulfate (phase II metabolism). A compound may also not be a useful drug because it is toxic, either because it modulates the target molecule too effectively or because it also binds to proteins other than the target. The liver and kidneys play central roles in drug metabolism and excretion.

36.2 Drug Candidates Can Be Discovered by Serendipity, Screening, or Design

Many drugs have been discovered by serendipity—that is, by chance observation. The antibiotic penicillin is produced by a mold that accidentally contaminated a culture dish, killing nearby bacteria. Drugs such as chlorpromazine and sildenafil were discovered to have beneficial, but unexpected, effects on human physiology. The cholesterol-lowering statin drugs were developed after large collections of compounds were screened for potentially interesting activities. Combinatorial chemistry methods have been developed to generate large collections of chemically related yet diverse compounds for screening. In some cases, the three-dimensional structure of a drug target is available and can be used to aid the design of potent and specific inhibitors. Examples of drugs designed in this manner are the HIV protease inhibitor indinavir and cyclooxygenase 2 inhibitors such as celecoxib.

36.3 Analyses of Genomes Hold Great Promise for Drug Discovery

The human genome encodes approximately 21,000 proteins, and many more if derivatives due to alternative mRNA splicing and post-translational modification are included. The genome sequences can be examined for potential drug targets. Large families of proteins known to participate in key physiological processes such as the protein kinases and 7TM receptors have each yielded several targets for which drugs have been developed. The genomes of model organisms also are useful for drug-development studies. Strains of mice with particular genes disrupted have been useful in validating certain drug targets. The genomes of bacteria, viruses, and parasites encode many potential drug targets that can be exploited owing to their important functions and their differences from human proteins, minimizing the potential for side effects. Genetic differences between individuals can be examined and correlated with differences in responses to drugs, potentially aiding both clinical treatments and drug development.

36.4 The Clinical Development of Drugs Proceeds Through Several Phases

Before compounds can be given to human beings as drugs, they must be extensively tested for safety and efficacy. Clinical trials are performed in stages: first testing safety, then safety and efficacy in a small population, and finally safety and efficacy in a larger population to detect rarer adverse effects. Largely due to the expenses associated with clinical trials, the cost of developing a new drug has been estimated to be more than $800 million. Even when a drug has been approved for use, complications can arise. With regard to infectious diseases and cancer, patients often develop resistance to a drug after it has been administered for a period of time because variants of the disease agent that are less susceptible to the drug arise and replicate, even when the drug is present.

KEY TERMS

ligand (p. 1035)
dissociation constant (K_d) (p. 1035)
side effect (p. 1035)
inhibition constant (K_i) (p. 1036)
Cheng–Prusoff equation (p. 1036)
ADME (p. 1036)
oral bioavailability (p. 1036)
Lipinski's rules (p. 1036)
compartment (p. 1037)
blood–brain barrier (p. 1037)
xenobiotic compounds (p. 1037)
drug metabolism (p. 1037)
oxidation (p. 1037)
conjugation (p. 1037)
phase I transformation (p. 1039)
phase II transformation (p. 1039)
first-pass metabolism (p. 1039)
glomeruli (p. 1039)
enterohepatic cycling (p. 1040)
therapeutic index (p. 1041)
atheroma (p. 1043)
myopathy (p. 1044)
high-throughput screening (p. 1045)
combinatorial chemistry (p. 1045)
split-pool synthesis (p. 1045)
structure-based drug design (p. 1046)
structure–activity relationship (SAR) (p. 1046)
placebo (p. 1052)
placebo effect (p. 1052)

PROBLEMS

1. *Routes to discovery.* For each of the following drugs, indicate whether the physiological effects of the drug were known before or after the target was identified.

(a) Penicillin
(b) Sildenafil (Viagra)
(c) Rofecoxib (Vioxx)
(d) Atorvastatin (Lipitor)
(e) Aspirin
(f) Indinavir (Crixivan)

2. *Lipinski's rules.* Which of the following compounds satisfy all of Lipinski's rules? [Log(P) values are given in parentheses.]

(a) Atenolol (0.23)
(b) Sildenafil (3.18)
(c) Indinavir (2.78)

3. *Calculating log tables.* Considerable effort has been expended to develop computer programs that can estimate log(P) values entirely on the basis of chemical structure. Why would such programs be useful?

4. *An ounce of prevention.* Legislation has been proposed that would require that N-acetylcysteine be added to acetaminophen tablets. Speculate about the role of this additive.

5. *Clinical-trial design.* Distinguish between Phase I and Phase II clinical trials in regard to number of persons enrolled, the state of health of the subjects, and the goals of the study.

6. *Drug interactions.* As noted in this chapter, coumadin can be a very dangerous drug because too much can cause uncontrolled bleeding. Persons taking coumadin must be careful about taking other drugs, particularly those that bind to albumin. Propose a mechanism for this drug–drug interaction.

7. *A bad combination.* Explain why drugs that inhibit P450 enzymes may be particularly dangerous when used in combination with other medications.

8. *Mechanistically speaking.* Name one advantage of a noncompetitive inhibitor as a potential drug compared with a competitive inhibitor.

9. *A helping hand.* You have developed a drug that is capable of inhibiting the ABC transporter MDR. Suggest a possible application for this drug in cancer chemotherapy.

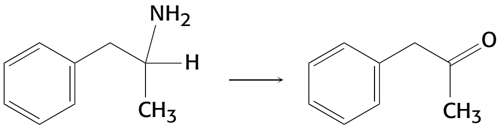
10. *Find the target.* Trypanosomes are unicellular parasites that cause sleeping sickness. During one stage of their life cycle, these organisms live in the bloodstream and derive all of their energy from glycolysis, which takes place in a specialized organelle called a glycosome inside the parasite. Propose potential targets for treating sleeping sickness. What are some potential difficulties with your approach?

11. *Knowledge is power.* How might genomic information be helpful for the effective use of imatinib mesylate (Gleevec) in cancer chemotherapy?

12. *Multiple targets, same goal.* Sildenafil induces its physiological effects by increasing the intracellular concentrations of cGMP, leading to muscle relaxation. On the basis of the scheme shown in Figure 36.17, identify another approach for increasing cGMP levels with a small molecule.

Mechanism Problem

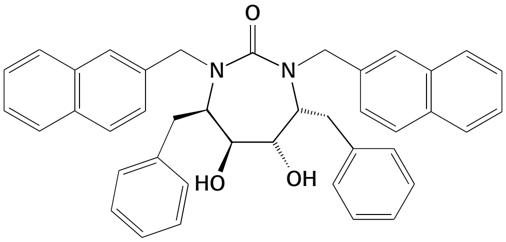
13. *Variations on a theme.* The metabolism of amphetamine by cytochrome P450 enzymes results in the conversion shown here. Propose a mechanism and indicate any additional products.



Amphetamine

Data Interpretation Problem

14. *HIV protease inhibitor design.* Compound A is one of a series that were designed to be potent inhibitors of HIV protease.



Compound A

Compound A was tested by using two assays: (1) direct inhibition of HIV protease in vitro and (2) inhibition of viral RNA production in HIV-infected cells, a measure of viral replication. The results of these assays are shown below. The HIV protease activity is measured with a substrate peptide present at a concentration equal to its K_M value.

Compound A (nM)	HIV protease activity (arbitrary units)
0	11.2
0.2	9.9
0.4	7.4
0.6	5.6
0.8	4.8
1	4.0
2	2.2
10	0.9
100	0.2

Compound A (nM)	Viral RNA production (arbitrary units)
0	760
1.0	740
2.0	380
3.0	280
4.0	180
5.0	100
10	30
50	20

Estimate the values for the K_I of compound A in the protease-activity assay and for its IC$_{50}$ in the viral-RNA-production assay.

Treating rats with the relatively high oral dose of 20 mg kg^{-1} results in a maximum concentration of compound A of 0.4 μM. On the basis of this value, do you expect compound A to be effective in preventing HIV replication when taken orally?

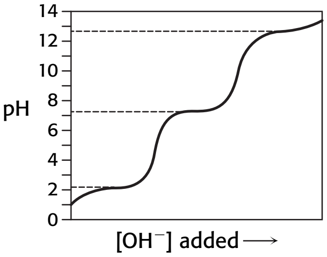
ANSWERS TO PROBLEMS

Chapter 1

1. The hydrogen-bond donors are the NH and NH_2 groups. The hydrogen-bond acceptors are the carbonyl oxygen atoms and those ring nitrogen atoms that are not bonded to hydrogen or to deoxyribose.
2. Interchange the positions of the single and double bonds in the six-membered ring.
3. (a) Ionic interactions; (b) van der Waals interactions.
4. Processes *a* and *b*.
5. $\Delta S_{\text{system}} = -661 \text{ J mol}^{-1} \text{ K}^{-1} (-158 \text{ kcal mol}^{-1} \text{ K}^{-1})$

$\Delta S_{\text{surroundings}} = +842 \text{ J mol}^{-1} \text{ K}^{-1} (+201 \text{ cal mol}^{-1} \text{ K}^{-1})$

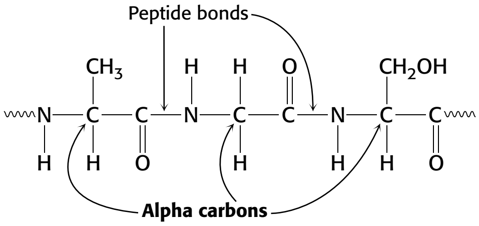
6. (a) 1.0; (b) 13.0; (c) 1.3; (d) 12.7
7. 2.88
8. 1.96
9. 55.5 M
10. 11.83
11. 447; 0.00050
12. 0.00066 M
13. 6.0
14. 5.53
15. 6.48
16. 7.8
17. 100
18. (a) 1.6; (b) 0.51; (c) 0.16.
19.



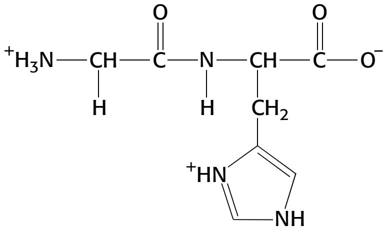
20. 0.1 M sodium acetate solution: 6.34; 6.03; 5.70; 4.75. 0.01 M sodium acetate solution: 5.90; 4.75; 3.38; 1.40.
21. 90 mM acetic acid; 160 mM sodium acetate, 0.18 moles acetic acid; 0.32 moles sodium acetate; 10.81 g acetic acid; 26.25 g sodium acetate.
22. 0.50 moles of acetic acid; 0.32 moles of NaOH; 30.03 g of acetic acid; 12.80 g of NaOH.
23. 250 mM; yes; no, it will also contain 90 mM NaCl.
24. 8.63 g Na_2HPO_4; 4.71 g NaH_2PO_4
25. 7.0; this buffer will not be very useful, because the pH value is far from the pK_a value.
26. $+1.45 \text{ kJ mol}^{-1} (+0.35 \text{ kcal mol}^{-1})$; $+57.9 \text{ kJ mol}^{-1} (+13.8 \text{ kcal mol}^{-1})$
27. There will be approximately 15 million differences.
28. (20!)/(10!*10!) = 184,756
29. 7.9%

Chapter 2

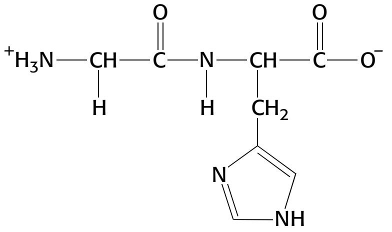
1. (A) Proline, Pro, P; (B) tyrosine, Tyr, Y; (C) leucine, Leu, L; (D) lysine, Lys, K.
2. (a) C, B, A; (b) D; (c) D, B; (d) B, D; (e) B.
3. (a) 6; (b) 2; (c) 3; (d) 1; (e) 4; (f) 5.
4. (a) Ala; (b) Tyr; (c) Ser; (d) His.
5. Ser, Glu, Tyr, Thr
6. (a) Alanine-glycine-serine; (b) Alanine; (c and d):



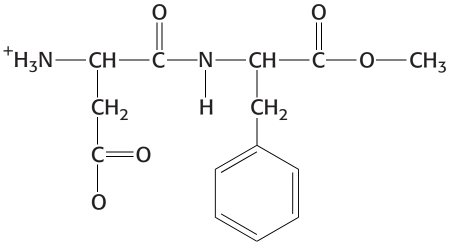
7. At pH 5.5, the net charge is +1:



At pH 7.5, the net charge is 0:



8. There are 20 choices for each of the 50 amino acids: 20^{50}, or 1×10^{65}.
9.



10. The (nitrogen–α carbon–carbonyl carbon) repeating unit.
11. Side chain is the functional group attached to the α-carbon atom of an amino acid.
12. Amino acid composition refers simply to the amino acids that make up the protein. The order is not specified. Amino acid sequence is the same as the primary structure—the sequence of amino acids from the amino terminal to the carboxyl terminal of the protein. Different proteins may have the same amino acid composition, but amino acid sequence identifies a unique protein.
13. (a) Each strand is 35 kDa and hence has about 318 residues (the mean residue mass is 110 daltons). Because the rise per residue in an α helix is 1.5 Å, the length is 477 Å. More precisely, for an α-helical coiled coil, the rise per residue is 1.46 Å; so the length is 464 Å. (b) Eighteen residues in each strand (40 minus 4 divided by 2) are in a β-sheet conformation. Because the rise per residue is 3.5 Å, the length is 63 Å.
14. The methyl group attached to the β-carbon atom of isoleucine sterically interferes with α-helix formation. In leucine, this methyl group is attached to the γ-carbon atom, which is farther from the main chain and hence does not interfere.

15. Proline and glycine. The cyclic side chain of proline linking the nitrogen and α-carbon atoms limits ϕ to a very narrow range (around -60 degrees). The lack of steric hindrance exhibited by the side chain hydrogen atom of glycine enables this amino acid to access a much greater area of the Ramachandran plot.

16. The first mutation destroys activity because valine occupies more space than alanine does, and so the protein must take a different shape, assuming that this residue lies in the closely packed interior. The second mutation restores activity because of a compensatory reduction of volume; glycine is smaller than isoleucine.

17. Loops invariably are on the surface of proteins, exposed to the environment. Because many proteins exist in aqueous environments, the exposed loops will be hydrophilic to interact with water.

18. The native conformation of insulin is not the thermodynamically most stable form, because it contains two separate chains linked by disulfide bonds. Insulin is formed from proinsulin, a single-chain precursor, that is cleaved to form insulin, a 51-residue molecule, after the disulfide bonds have formed.

19. A segment of the main chain of the protease could hydrogen bond to the main chain of the substrate to form an extended parallel or antiparallel pair of β strands.

20. Glycine has the smallest side chain of any amino acid. Its size is often critical in allowing polypeptide chains to make tight turns or to approach one another closely.

21. Glutamate, aspartate, and the terminal carboxylate can form salt bridges with the guanidinium group of arginine. In addition, this group can be a hydrogen-bond donor to the side chains of glutamine, asparagine, serine, threonine, aspartate, tyrosine, and glutamate and to the main-chain carbonyl group. Histidine can form hydrogen bonds with arginine at pH 7.

22. Disulfide bonds in hair are broken by adding a thiol-containing reagent and applying gentle heat. The hair is curled, and an oxidizing agent is added to re-form disulfide bonds to stabilize the desired shape.

23. Some proteins that span biological membranes are "the exceptions that prove the rule" because they have the reverse distribution of hydrophobic and hydrophilic amino acids. For example, consider *porins,* proteins found in the outer membranes of many bacteria. Membranes are built largely of hydrophobic chains. Thus, porins are covered on the outside largely with hydrophobic residues that interact with the neighboring hydrophobic chains. In contrast, the center of the protein contains many charged and polar amino acids that surround a water-filled channel running through the middle of the protein. Thus, because porins function in hydrophobic environments, they are "inside out" relative to proteins that function in aqueous solution.

24. The amino acids would be hydrophobic in nature. An α helix is especially suited to crossing a membrane because all of the amide hydrogen atoms and carbonyl oxygen atoms of the peptide backbone take part in intrachain hydrogen bonds, thus stabilizing these polar atoms in a hydrophobic environment.

25. This example demonstrates that the $\mathrm{p}K_\mathrm{a}$ values are affected by the environment. A given amino acid can have a variety of $\mathrm{p}K_\mathrm{a}$ values, depending on the chemical environment inside the protein.

26. Recall that hemoglobin exists as a tetramer while myoglobin is a monomer. Consequently, the hydrophobic residues on the surface of hemoglobin subunits are probably involved in van der Waals interactions with similar regions on the other subunits, and will be shielded from the aqueous environment by this interaction.

27. A possible explanation is that the severity of the symptoms corresponds to the degree of structural disruption. Hence, substitution of alanine for glycine might result in mild symptoms, but substitution of the much larger tryptophan may result in little or no collagen triple-helix formation.

28. The energy barrier that must be crossed to go from the polymerized state to the hydrolyzed state is large even though the reaction is thermodynamically favorable.

29. Using the Henderson–Hasselbach equation, we find the ratio of alanine-COOH to alanine-$\mathrm{COO^-}$ at pH 7 to be 10^{-4}. The ratio of alanine-$\mathrm{NH_2}$ to alanine-$\mathrm{NH_3^+}$, determined in the same fashion, is 10^{-1}. Thus, the ratio of neutral alanine to the zwitterionic species is $10^{-4} \times 10^{-1} = 10^{-5}$.

30. The assignment of absolute configuration requires the assignment of priorities to the four groups connected to a tetrahedral carbon atom. For all amino acids except cysteine, the priorities are: (1) amino group; (2) carbonyl group; (3) side chain; (4) hydrogen. For cysteine, because of the sulfur atom in its side chain, the side chain has a greater priority than does the carbonyl group, leading to the assignment of an *R* rather than *S* configuration.

31. ELVISISLIVINGINLASVEGAS

32. No, Pro–X would have the characteristics of any other peptide bond. The steric hindrance in X–Pro arises because the R group of Pro is bonded to the amino group. Hence, in X–Pro, the proline R group is near the R group of X, which would not be the case in Pro–X.

33. A, c; B, e; C, d; D, a; E, b.

34. The reason is that the wrong disulfides formed pairs in urea. There are 105 different ways of pairing eight cysteine molecules to form four disulfides; only one of these combinations is enzymatically active. The 104 wrong pairings have been picturesquely termed "scrambled" ribonuclease.

Chapter 3

1. (a) Phenyl isothiocyanate; (b) urea; β-mercaptoethanol to reduce disulfides; (c) chymotrypsin; (d) CNBr; (e) trypsin.

2. For each cell within an organism, the genome is a fixed property. However, the proteome is dynamic, reflecting different environmental conditions and external stimuli. Two different cell types will likely express different subsets of proteins encoded within the genome.

3. The *S*-aminoethylcysteine side chain resembles that of lysine. The only difference is a sulfur atom in place of a methylene group.

4. A 1 mg $\mathrm{ml^{-1}}$ solution of myoglobin (17.8 kDa; Table 3.2) corresponds to 5.62×10^{-5} M. The absorbance of a 1-cm path length is 0.84, which corresponds to an I_0/I ratio of 6.96. Hence 14.4% of the incident light is transmitted.

5. The sample was diluted 1000-fold. The concentration after dialysis is thus 0.001 M, or 1 mM. You could reduce the salt concentration by dialyzing your sample, now 1 mM, in more buffer free of $\mathrm{(NH_4)_2SO_4}$.

6. If the salt concentration becomes too high, the salt ions interact with the water molecules. Eventually, there will not be enough water molecules to interact with the protein, and the protein will precipitate. If there is lack of salt in a protein solution, the proteins may interact with one another—the positive charges on one protein with the negative charges on another or several others. Such an aggregate becomes too large to be solubilized by water alone. If salt is added, the salt neutralizes the charges on the proteins, preventing protein–protein interactions.

7. Tropomyosin is rod shaped, whereas hemoglobin is approximately spherical.

8. The frictional coefficient, f, and the mass, m, determine s. Specifically, f is proportional to r (see equation 2 on p. 71). Hence, f is proportional to $m^{1/3}$, and so s is proportional to $m^{2/3}$ (see the

equation on p. 76). An 80-kDa spherical protein undergoes sedimentation 1.59 times as rapidly as a 40-kDa spherical protein.

9. The long hydrophobic tail on the SDS molecule (see p. 72) disrupts the hydrophobic interactions in the interior of the protein. The protein unfolds, with the hydrophobic R groups now interacting with SDS rather than with one another.

10. 50 kDa

11. The protein may be modified. For instance, asparagine residues in the protein may be modified with carbohydrate units (Section 2.6).

12. A fluorescence-labeled derivative of a bacterial degradation product (e.g., a formylmethionyl peptide) would bind to cells containing the receptor of interest.

13. (a) Trypsin cleaves after arginine (R) and lysine (K), generating AVGWR, VK, and S. Because they differ in size, these products could be separated by molecular exclusion chromatography. (b) Chymotrypsin, which cleaves after large aliphatic or aromatic R groups, generates two peptides of equal size (AVGW) and (RVKS). Separation based on size would not be effective. The peptide RVKS has two positive charges (R and K), whereas the other peptide is neutral. Therefore, the two products could be separated by ion-exchange chromatography.

14. Antibody molecules bound to a solid support can be used for affinity purification of proteins for which a ligand molecule is not known or unavailable.

15. If the product of the enzyme-catalyzed reaction is highly antigenic, it may be possible to obtain antibodies to this particular molecule. These antibodies can be used to detect the presence of product by ELISA, providing an assay format suitable for the purification of this enzyme.

16. An inhibitor of the enzyme being purified might have been present and subsequently removed by a purification step. This removal would lead to an apparent increase in the total amount of enzyme present.

17. Many proteins have similar masses but different sequences and different patterns when digested with trypsin. The set of masses of tryptic peptides forms a detailed "fingerprint" of a protein that is very unlikely to appear at random in other proteins regardless of size. (A conceivable analogy is: "Just as similarly sized fingers will give different individual fingerprints, so also similarly sized proteins will give different digestion patterns with trypsin.")

18. Isoleucine and leucine are isomers and, hence, have identical masses. Peptide sequencing by mass spectrometry as described in this chapter is incapable of distinguishing these residues. Further analytical techniques are required to differentiate these residues.

19. See the table at the bottom of the page.

20. (a) Ion exchange chromatography will remove Proteins A and D, which have substantially lower isoelectric point; then gel filtration chromatography will remove Protein C, which has a lower molecular weight. (b) If Protein B carries a His tag, a single affinity chromatography step with an immobilized nickel(II) column may be sufficient to isolate the desired protein from the others.

21. Protein crystal formation requires the ordered arrangement of identically positioned molecules. Proteins with flexible linkers can introduce disorder into this arrangement and prevent the formation of suitable crystals. A ligand or binding partner may induce an ordered conformation to this linker and could be included in the solution to facilitate crystal growth. Alternatively, the individual domains separated by the linker may be expressed by recombinant methods and their crystal structures solved separately.

22. Treatment with urea will disrupt noncovalent bonds. Thus the original 60-kDa protein must be made of two 30-kDa subunits. When these subunits are treated with urea and β-mercaptoethanol, a single 15-kDa species results, suggesting that disulfide bonds link the 30-kDa subunits.

23. (a) Electrostatic repulsion between positively charged ε-amino groups hinders α-helix formation at pH 7. At pH 10, the side chains become deprotonated, allowing α-helix formation.
(b) Poly-L-glutamate is a random coil at pH 7 and becomes α helical below pH 4.5 because the γ-carboxylate groups become protonated.

24. The difference between the predicted and the observed masses for this fragment equals 28.0, exactly the mass shift that would be expected in a formylated peptide. This peptide is likely formylated at its amino terminus, and corresponds to the most N-terminal fragment of the protein.

25. Light was used to direct the synthesis of these peptides. Each amino acid added to the solid support contained a photolabile protecting group instead of a *t*-Boc protecting group at its α-amino group. Illumination of selected regions of the solid support led to the release of the protecting group, which exposed the amino groups in these sites to make them reactive. The pattern of masks used in these illuminations and the sequence of reactants define the ultimate products and their locations.

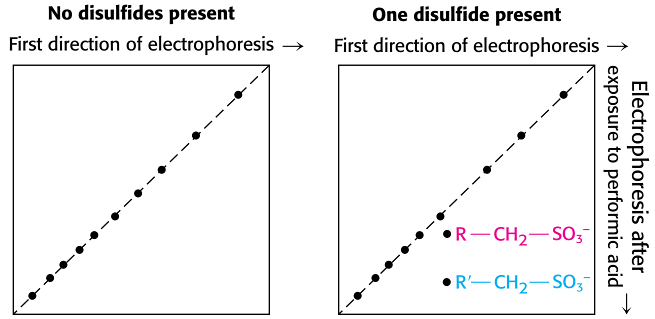
26. Mass spectrometry is highly sensitive and capable of detecting the mass difference between a protein and its deuterated counterpart. Fragmentation techniques can be used to identify the amino acids that retained the isotope label. Alternatively, NMR spectroscopy can be used to detect the isotopically labeled atoms because the deuteron and the proton have very different nuclear-spin properties.

27. First amino acid: A
Last amino acid: R (not cleaved by carboxypeptidase).
Sequence of N-terminal tryptic peptide: AVR (tryptic peptide ends in K)
Sequence of N-terminal chymotryptic peptide: AVRY (chymotryptic peptide ends in Y)
Sequence: AVRYSR

28. First amino acid: S
Last amino acid: L
Cyanogen bromide cleavage: M is 10th position
C-terminal residues are: (2S,L,W)
Amino-terminal residues: (G,K,S,Y), tryptic peptide, ends in K
Amino-terminal sequence: SYGK
Chymotryptic peptide order: (S,Y), (G,K,L), (F,I,S), (M,T), (S,W), (S,L)
Sequence: SYGKLSIFTMSWSL

Purification procedure	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Purification level	Yield (%)
Crude extract	20,000	4,000,000	200	1	100
$(NH_4)_2SO_4$ precipitation	5,000	3,000,000	600	3	75
DEAE–cellulose chromatography	1,500	1,000,000	667	3.3	25
Gel-filtration chromatography	500	750,000	1,500	7.5	19
Affinity chromatography	45	675,000	15,000	75	17

29. If the protein did not contain any disulfide bonds, then the electrophoretic mobility of the trypsin fragments would be the same before and after performic acid treatment: all the fragments would lie along the diagonal of the paper. If one disulfide bond were present, the disulfide-linked trypsin fragments would run as a single peak in the first direction, then would run as two separate peaks after performic acid treatment. The result would be two peaks appearing off the diagonal:


No disulfides present
First direction of electrophoresis →
One disulfide present
First direction of electrophoresis →
Electrophoresis after exposure to performic acid ↓
R—CH₂—SO₃⁻
R′—CH₂—SO₃⁻

These fragments could then be isolated from the chromatography paper and analyzed by mass spectrometry to determine their amino acid composition and thus identify the cysteines participating in the disulfide bond.

Chapter 4

1. A nucleoside is a base attached to a ribose or deoxyribose sugar. A nucleotide is a nucleoside with one or more phosphoryl groups attached to the ribose or deoxyribose.
2. Hydrogen-bond pairing between the base A and the base T as well as hydrogen-bond pairing between the base G and the base C in DNA.
3. T is always equal to A, and so these two nucleotides constitute 40% of the bases. G is always equal to C, and so the remaining 60% must be 30% G and 30% C.
4. Nothing, because the base-pair rules do not apply to single-stranded nucleic acids.
5. (a) TTGATC; (b) GTTCGA; (c) ACGCGT; (d) ATGGTA.
6. (a) $[T] + [C] = 0.46$. (b) $[T] = 0.30$, $[C] = 0.24$, and $[A] + [G] = 0.46$.
7. Stable hydrogen bonding occurs only between GC and AT pairs. Moreover, two purines are too large to fit inside the double helix, and two pyrimidines are too small to form base pairs with each other.
8. The thermal energy causes the chains to wiggle about, which disrupts the hydrogen bonds between base pairs and the stacking forces between bases and thereby causes the strands to separate.
9. The probability that any sequence will appear is $1/4^n$, where 4 is the number of nucleotides and n is the length of the sequence. The probability of any 15-base sequence appearing is $1/4^{15}$, or 1/1,073,741,824. Thus, a 15-nucleotide sequence would be likely to appear approximately three times (3 billion × probability of appearance). The probability of a 16-base sequence appearing is $1/4^{16}$, which is equal to 1/4,294,967,296. Such a sequence will be unlikely to appear more than once.
10. One end of a nucleic acid polymer ends with a free 5′-hydroxyl group (or a phosphoryl group esterified to the hydroxyl group), and the other end has a free 3′-hydroxyl group. Thus, the ends are different. Two strands of DNA can form a double helix only if the strands are running in different directions—that is, have opposite polarity.
11. Although the individual bonds are weak, the population of thousands to millions of such bonds provides much stability. There is strength in numbers.
12. There would be too much charge repulsion from the negative charges on the phosphoryl groups. These charges must be countered by the addition of cations.
13. The three forms are the A-DNA, the B-DNA and the Z-DNA, with B-DNA being the most common. There are many differences (Table 4.2). Some key differences are: A-DNA and B-DNA are right-handed, whereas Z-DNA is left-handed. A-DNA forms in less-hydrated conditions than does B-DNA. The A form is shorter and wider than the B form.
14. 5.88×10^3 base pairs.
15. The diameter of DNA is 20 Å and 1 Å = 0.1 nm, so the diameter is 2 nm. Because 1 μm = 10^3 nm, the length is 2×10^4 nm. Thus, the axial ratio is 1×10^4.
16. A template is the sequence of DNA or RNA that directs the synthesis of a complementary sequence. A primer is the initial segment of a polymer that is to be extended during elongation.
17. In conservative replication, after 1.0 generation, half of the molecules would be ^{15}N-^{15}N, the other half ^{14}N-^{14}N. After 2.0 generations, one-quarter of the molecules would be ^{15}N-^{15}N, the other three-quarters ^{14}N-^{14}N. Hybrid ^{14}N-^{15}N molecules would not be observed in conservative replication.
18. The nucleotides used for DNA synthesis have the triphosphate attached to the 5′-hydroxyl group with free 3′-hydroxyl groups. Such nucleotides can be utilized only for 5′-to-3′ DNA synthesis.
19. (a) Tritiated thymine or tritiated thymidine. (b) dATP, dGTP, dCTP, and TTP labeled with ^{32}P in the innermost (α) phosphorus atom.
20. Molecules in parts *a* and *b* would not lead to DNA synthesis, because they lack a 3′-OH group (a primer). The molecule in part *d* has a free 3′-OH group at one end of each strand but no template strand beyond. Only the molecule in part *c* would lead to DNA synthesis.
21. A retrovirus is a virus that has RNA as its genetic material. However, for the information to be expressed, it must first be converted into DNA, a reaction catalyzed by the enzyme reverse transcriptase. Thus, at least initially, information flow is opposite that of a normal cell: RNA → DNA rather than DNA → RNA.
22. A thymidylate oligonucleotide should be used as the primer. The poly(A) template specifies the incorporation of T; hence, radioactive thymidine triphosphate (labeled in the α phosphoryl group) should be used in the assay.
23. The ribonuclease serves to degrade the RNA strand, a necessary step in forming duplex DNA from the RNA–DNA hybrid.
24. Treat one aliquot of the sample with ribonuclease and another with deoxyribonuclease. Test these nuclease-treated samples for infectivity.
25. Deamination changes the original G · C base pair into a G · U pair. After one round of replication, one daughter duplex will contain a G · C pair and the other duplex will contain an A · U pair. After two rounds of replication, there will be two G · C pairs, one A · U pair, and one A · T pair.
26. (a) $4^8 = 65{,}536$. In computer terminology, there are 64K 8-mers of DNA.
(b) A bit specifies two bases (say, A and C) and a second bit specifies the other two (G and T). Hence, two bits are needed to specify a single nucleotide (base pair) in DNA. For example, 00, 01, 10, and 11 could encode A, C, G, and T. An 8-mer stores 16 bits ($2^{16} = 65{,}536$), the *E. coli* genome (4.6×10^6 bp) stores 9.2×10^6 bits, and the human genome (3.0×10^9 bases) stores 6.0×10^9 bits of genetic information.
(c) A standard CD can hold about 700 megabytes, which is equal to 5.6×10^9 bits. A large number of 8-mer sequences could be

stored on such a CD. The DNA sequence of *E. coli* could be written on a single CD with room to spare for a lot of music. One CD would not be quite enough to record the entire human genome.

27. (a) Deoxyribonucleoside triphosphates versus ribonucleoside triphosphates.
(b) 5′ → 3′ for both.
(c) Semiconserved for DNA polymerase I; conserved for RNA polymerase.
(d) DNA polymerase I needs a primer, whereas RNA polymerase does not.

28. The template strand has a sequence complementary to that of the RNA transcript. The coding strand has the same sequence as that of the RNA transcript except for thymine (T) in place of uracil (U).

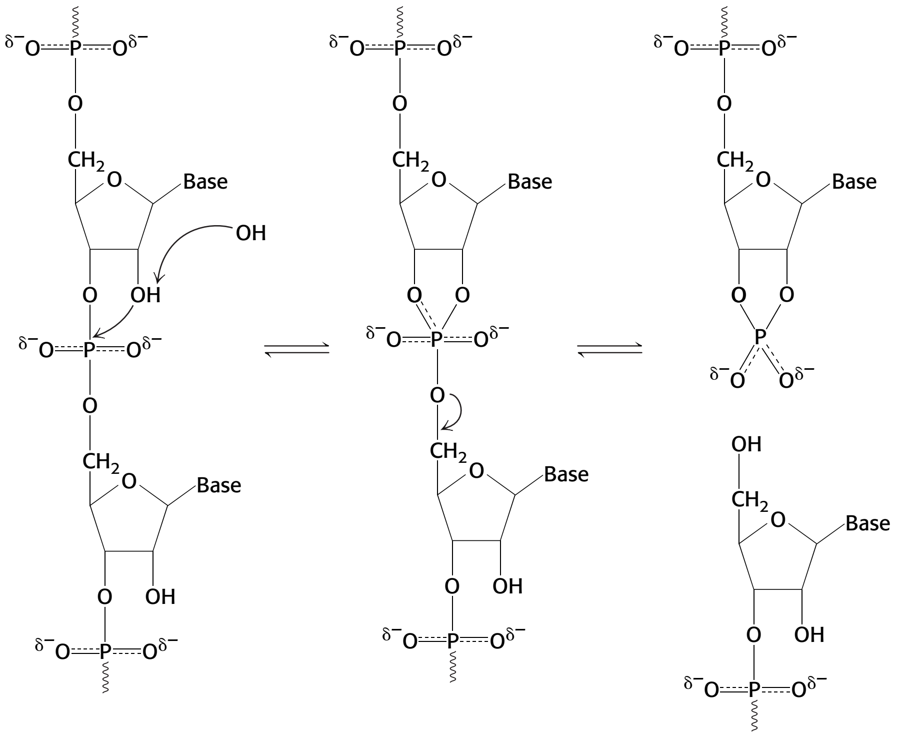
29. Messenger RNA encodes the information that, on translation, yields a protein. Ribosomal RNA is the catalytic component of ribosomes, the molecular complexes that synthesize proteins. Transfer RNA is an adaptor molecule, capable of binding a specific amino acid and recognizing a corresponding codon. Transfer RNAs with attached amino acids are substrates for the ribosome.

30. Three nucleotides encode an amino acid; the code is nonoverlapping; the code has no punctuation; the code exhibits directionality; the code is degenerate.

31. (a) 5′-UAACGGUACGAU-3′
(b) Leu-Pro-Ser-Asp-Trp-Met
(c) Poly(Leu-Leu-Thr-Tyr)

32. The 2′-OH group in RNA acts as an intramolecular nucleophile. In the alkaline hydrolysis of RNA, it forms a 2′-3′ cyclic intermediate.

33.



34. Gene expression is the process of expressing the information of a gene in its functional molecular form. For many genes, the functional information is a protein molecule. Thus, gene expression includes transcription and translation.

35. A nucleotide sequence whose bases represent the most-common, but not necessarily the only, members of the sequence. A consensus sequence can be thought of as the average of many similar sequences.

36. Cordycepin terminates RNA synthesis. An RNA chain containing cordycepin lacks a 3′-OH group.

37. Only single-stranded RNA can serve as a template for protein synthesis.

38. Degeneracy of the code refers to the fact that most amino acids are encoded by more than one codon.

39. If only 20 of the 64 possible codons encoded amino acids, then a mutation that changed a codon would likely result in a nonsense codon, leading to termination of protein synthesis. With degeneracy, a nucleotide change might yield a synonym or a codon for an amino acid with similar chemical properties.

40. (a) 2, 4, 8; (b) 1, 6, 10; (c) 3, 5, 7, 9.

41. (a) 3; (b) 6; (c) 2; (d) 5; (e) 7; (f) 1; (g) 4.

42. Incubation with RNA polymerase and only UTP, ATP, and CTP led to the synthesis of only poly(UAC). Only poly(GUA) was formed when GTP was used in place of CTP.

43. A peptide terminating with Lys (UGA is a stop codon), another containing -Asn-Glu-, and a third containing -Met-Arg-.

44. Phe-Cys-His-Val-Ala-Ala

45. Exon shuffling is a molecular process that can lead to the generation of new proteins by the rearrangement of exons within genes. Because many exons encode functional protein domains, exon shuffling is a rapid and efficient means of generating new genes.

46. Alternate splicing allows one gene to code for several different but related proteins.

47. It shows that the genetic code and the biochemical means of interpreting the code are common to even very distantly related life forms. It also testifies to the unity of life: that all life arose from a common ancestor.

48. (a) A codon for lysine cannot be changed to one for aspartate by the mutation of a single nucleotide. (b) Arg, Asn, Gln, Glu, Ile, Met, or Thr.

49. The genetic code is degenerate. Of the 20 amino acids, 18 are specified by more than one codon. Hence, many nucleotide changes (especially in the third base of a codon) do not alter the nature of the encoded amino acid. Mutations leading to an altered amino acid are usually more deleterious than those that do not and hence are subject to more stringent selection.

50. GC base pairs have three hydrogen bonds compared with two for AT base pairs. Thus, the higher content of GC means more hydrogen bonds and greater helix stability.

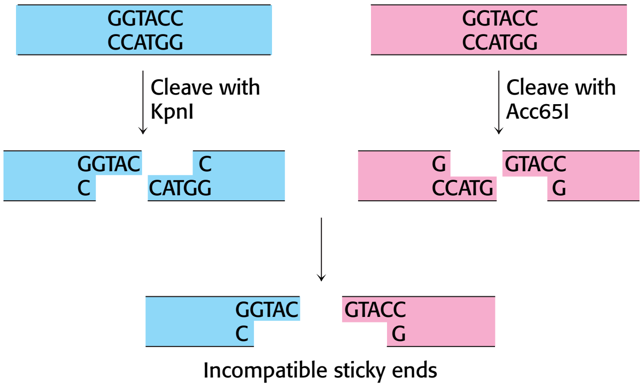
51. C_0t value essentially corresponds to the complexity of the DNA sequence—in other words, how long it will take for a sequence of DNA to find its complementary strand to form a double helix. The more complex the DNA, the slower it reassociates to make the double-stranded form.

52. Increasing salt increases the melting temperature. Because the DNA backbone is negatively charged, there is a tendency for charge repulsion to destabilize the helix and cause it to melt. The addition of salt neutralizes the charge repulsion, thereby stabilizing

the double helix. The results show that within the parameters of the experiment, more salt results in more stabilization, which is reflected as a higher melting temperature.

Chapter 5

1. *Taq* polymerase is the DNA polymerase from the thermophilic bacterium that lives in hot springs. Consequently, it is heat stable and can withstand the high temperatures required for PCR without denaturing.
2. Ovalbumin cDNA should be used. *E. coli* lacks the machinery to splice the primary transcript arising from genomic DNA.
3. Consistent with its planar, aromatic structure, ethidium bromide is a DNA intercalator: it aligns itself between the paired bases in a DNA duplex.
4. The presence of the AluI sequence would, on average, be $(1/4)^4$, or 1/256, because the likelihood of any base being at any position is one-fourth and there are four positions. By the same reasoning, the presence of the NotI sequence would be $(1/4)^8$, or 1/65,536. Thus, the average product of digestion by AluI would be 250 base pairs (0.25 kb) in length, whereas that by NotI would be 66,000 base pairs (66 kb) in length.
5. No, because most human genes are much longer than 4 kb. A fragment would contain only a small part of a complete gene.
6. Southern blotting of an MstII digest would distinguish between the normal and the mutant genes. The loss of a restriction site would lead to the replacement of two fragments on the Southern blot by a single longer fragment. Such a finding would not prove that GTG replaced GAG; other sequence changes at the restriction site could yield the same result.
7. Although the two enzymes cleave the same recognition site, they each break different bonds within the 6-bp sequence. Cleavage by KpnI yields an overhang on the $3'$ strand, whereas cleavage by Acc65I produces an overhang on the $5'$ strand. These sticky ends do not overlap.



8. A simple strategy for generating many mutants is to synthesize a degenerate set of cassettes by using a mixture of activated nucleosides in particular rounds of oligonucleotide synthesis. Suppose that the 30-bp coding region begins with GTT, which encodes valine. If a mixture of all four nucleotides is used in the first and second rounds of synthesis, the resulting oligonucleotides will begin with the sequence XYT (where X and Y denote A, C, G, or T). These 16 different versions of the cassette will encode proteins containing either Phe, Leu, Ile, Val, Ser, Pro, Thr, Ala, Tyr, His, Asn, Asp, Cys, Arg, or Gly at the first position. Likewise, degenerate cassettes can be made in which two or more codons are simultaneously varied.
9. Because PCR can amplify as little as one molecule of DNA, statements claiming the isolation of ancient DNA need to be greeted with some skepticism. The DNA would need to be sequenced. Is it similar to human, bacterial, or fungal DNA? If so, contamination is the likely source of the amplified DNA. Is it similar to that of birds or crocodiles? This sequence similarity would strengthen the case that it is dinosaur DNA because these species are evolutionarily close to dinosaurs.
10. PCR amplification is greatly hindered by the presence of G–C-rich regions within the template. Owing to their high melting temperatures, these templates do not denature easily, preventing the initiation of an amplification cycle. In addition, rigid secondary structures prevent the progress of DNA polymerase along the template strand during elongation.
11. At high temperatures of hybridization, only very close matches between primer and target would be stable because all (or most) of the bases would need to find partners to stabilize the primer–target helix. As the temperature is lowered, more mismatches would be tolerated; so the amplification is likely to yield genes with less sequence similarity. In regard to the yeast gene, synthesize primers corresponding to the ends of the gene, and then use these primers and human DNA as the target. If nothing is amplified at 54°C, the human gene differs from the yeast gene, but a counterpart may still be present. Repeat the experiment at a lower temperature of hybridization.
12. Digest genomic DNA with a restriction enzyme, and select the fragment that contains the known sequence. Circularize this fragment. Then carry out PCR with the use of a pair of primers that serve as templates for the synthesis of DNA away from the known sequence.
13. The encoded protein contains four repeats of a specific sequence.
14. Use chemical synthesis or the polymerase chain reaction to prepare hybridization probes that are complementary to both ends of the known (previously isolated) DNA fragment. Challenge clones representing the library of DNA fragments with both of the hybridization probes. Select clones that hybridize to one of the probes but not the other; such clones are likely to represent DNA fragments that contain one end of the known fragment along with the adjacent region of the particular chromosome.
15. The codon(s) for each amino acid can be used to determine the number of possible nucleotide sequences that encode each peptide sequence (Table 4.5):

Ala–Met–Ser–Leu–Pro–Trp:
$4 \times 1 \times 6 \times 6 \times 4 \times 1 = 576$ total sequences

Gly–Trp–Asp–Met–His–Lys:
$4 \times 1 \times 2 \times 1 \times 2 \times 2 = 32$ total sequences

Cys–Val–Trp–Asn–Lys–Ile:
$2 \times 4 \times 1 \times 2 \times 2 \times 3 = 96$ total sequences

Arg–Ser–Met–Leu–Gln–Asn:
$6 \times 6 \times 1 \times 6 \times 2 \times 2 = 864$ total sequences

The set of DNA sequences encoding the peptide Gly-Trp-Asp-Met-His-Lys would be most ideal for probe design because it encompasses only 32 total oligonucleotides.
16. Within a single species, individual dogs show enormous variation in body size and substantial diversity in other physical characteristics. Therefore, genomic analysis of individual dogs would provide valuable clues concerning the genes responsible for the diversity within the species.
17. On the basis of the comparative genome map shown in Figure 5.28, the region of greatest overlap with human chromosome 20 can be found on mouse chromosome 2.
18. T_m is the melting temperature of a double-stranded nucleic acid. If the melting temperatures of the primers are too different,

the extent of hybridization with the target DNA will differ during the annealing phase, which would result in differential replications of the strands.
19. Careful comparison of the sequences reveals that there is a 7-bp region of complementarity at the 3′ ends of these two primers:

5′-GGATCGATGCTCGCGA-3′
| | | | | | |
3′-GAGCGCTGGGCTAGGA-5′

In a PCR experiment, these primers would likely anneal to one another, preventing their interaction with the template DNA. During DNA synthesis by the polymerase, each primer would act as a template for the other primer, leading to the amplification of a 25-bp sequence corresponding to the overlapped primers.
20. A mutation in person B has altered one of the alleles for gene *X*, leaving the other intact. The fact that the mutated allele is smaller suggests that a deletion has occurred in one copy of the gene. The one functioning copy is transcribed and translated and apparently produces enough protein to render the person asymptomatic.

Person C has only the smaller version of the gene. This gene is neither transcribed (negative northern blot) nor translated (negative western blot).

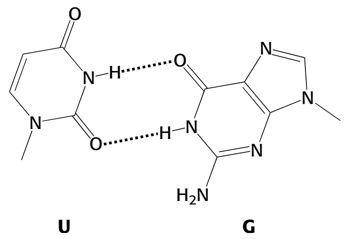
Person D has a normal-size copy of the gene but no corresponding RNA or protein. There may be a mutation in the promoter region of the gene that prevents transcription.

Person E has a normal-size copy of the gene that is transcribed, but no protein is made, which suggests that a mutation prevents translation. There are a number of possible explanations, including a mutation that introduced a premature stop codon in the mRNA.

Person F has a normal amount of protein but still displays the metabolic problem. This finding suggests that the mutation affects the activity of the protein—for instance, a mutation that compromises the active site of enzyme Y.
21. Chongqing: residue 2, L → R, CTG → CGG
Karachi: residue 5, A → P, GCC → CCC
Swan River: residue 6, D → G, GAC → GGC
22. This particular person is heterozygous for this particular mutation: one allele is wild type, whereas the other carries a point mutation at this position. Both alleles are PCR amplified in this experiment, yielding the "dual peak" appearance on the sequencing chromatogram.

Chapter 6

1. There are 26 identities and two gaps for a score of 210. The two sequences are approximately 26% identical. This level of homology is likely to be statistically significant.
2. They are likely related by divergent evolution, because three-dimensional structure is more conserved than is sequence identity.
3. (1) Identity score = −25; Blosum score = 14; (2) identity score = 15; Blosum score = 3.
4. U. One possible structure:



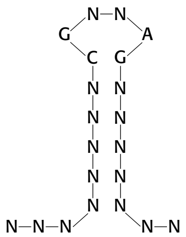
5. There are 4^{40}, or 1.2×10^{24}, different molecules. Each molecule has a mass of 2.2×10^{-20}, because 1 mol of polymer has a mass of 330 g $\text{mol}^{-1} \times 40$, and there are 6.02×10^{23} molecules per mole. Therefore, 26.4 kg of RNA would be required.
6. Because three-dimensional structure is much more closely associated with function than is sequence, tertiary structure is more evolutionarily conserved than is primary structure. In other words, protein function is the most important characteristic, and protein function is determined by structure. Thus, the structure must be conserved but not necessarily a specific amino acid sequence.
7. Alignment score of sequences (1) and (2) is $6 \times 10 = 60$. Many answers are possible, depending on the randomly reordered sequence. A possible result is

Shuffled sequence: (2) `TKADKAGEYL`

Alignment: (1) `ASNFLDKAGK`
(2) `TKADKAGEYL`

Alignment score is $4 \times 10 = 40$.

8. (a) Almost certainly diverged from a common ancestor. (b) Almost certainly diverged from a common ancestor. (c) May have diverged from a common ancestor, but the sequence alignment may not provide supporting evidence. (d) May have diverged from a common ancestor, but the sequence alignment is unlikely to provide supporting evidence.
9. Replacement of cysteine, glycine, and proline never yields a positive score. Each of these residues exhibits features unlike those of its other 19 counterparts: cysteine is the only amino acid capable of forming disulfide bonds, glycine is the only amino acid without a side-chain and is highly flexible, and proline is the only amino acid that is highly constrained through the bonding of its side chain to its amine nitrogen.
10. Protein A is clearly homologous to protein B, given 65% sequence identity, and so A and B are expected to have quite similar three-dimensional structures. Likewise, proteins B and C are clearly homologous, given 55% sequence identity, and so B and C are expected to have quite similar three-dimensional structures. Thus, proteins A and C are likely to have similar three-dimensional structures, even though they are only 15% identical in sequence.
11. The likely secondary structure is

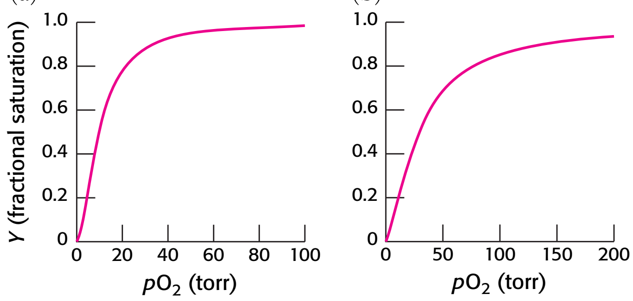


12. To detect pairs of residues with correlated mutations, there must be variability in these sequences. If the alignment is overrepresented by closely related organisms, there may not be enough changes in their sequences to allow the identification of potential base-pairing patterns.
13. After RNA molecules have been selected and reverse transcribed, PCR is performed to introduce additional mutations into these strands. The use of this error-prone, thermostable polymerase in the amplification step would enhance the efficiency of this random mutagenesis.
14. The initial pool of RNA molecules used in a molecular-evolution experiment is typically much smaller than the total number of possible sequences. Hence, the best possible RNA sequences will likely not be represented in the initial set of oligonucleotides.

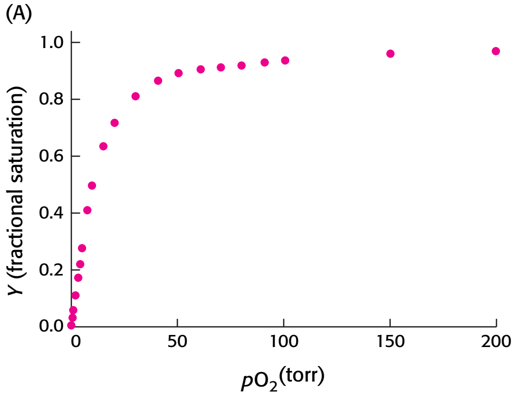
Mutagenesis of the initial selected RNA molecules allows for the iterative improvement of these sequences for the desired property.
15. 107 or 108 identities (depending on which annotated human sequence is chosen).

Chapter 7

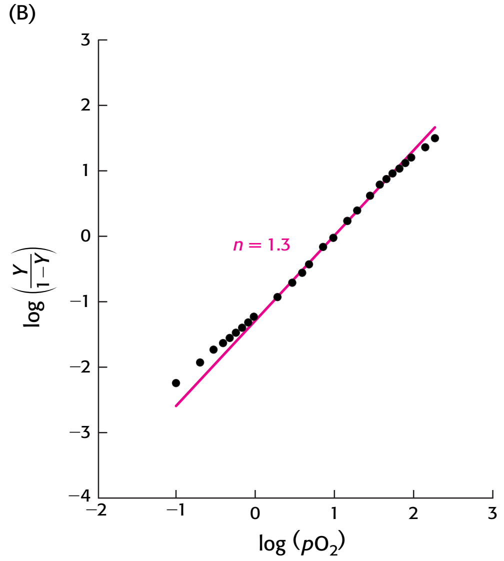
1. The whale swims long distances between breaths. A high concentration of myoglobin in the whale muscle maintains a ready supply of oxygen for the muscle between breathing episodes.
2. (a) 2.96×10^{-11} g
(b) 2.74×10^{8} molecules
(c) No. There would be 3.17×10^{8} hemoglobin molecules in a red cell if they were packed in a cubic crystalline array. Hence, the actual packing density is about 84% of the maximum possible.
3. 2.65 g (or 4.75×10^{-2} mol) of Fe
4. (a) In human beings, 1.44×10^{-2} g (4.49×10^{-4} mol) of O_2 per kilogram of muscle. In sperm whale, 0.144 g (4.49×10^{-3} mol) of O_2 per kilogram.
(b) 128
5. The pK_a is (a) lowered; (b) raised; and (c) raised.
6. Deoxy Hb A contains a complementary site, and so it can add on to a fiber of deoxy Hb S. The fiber cannot then grow further, because the terminal deoxy Hb A molecule lacks a sticky patch.
7. 62.7% oxygen-carrying capacity
8. Myoglobin does not exhibit a Bohr effect. The interactions responsible for mediating the Bohr effect in hemoglobin are dependent upon a tetrameric structure. Myoglobin is a monomer.
9. A higher concentration of BPG would shift the oxygen-binding curve to the right, causing an increase in P_{50}. The larger value of P_{50} would promote dissociation of oxygen in the tissues and would thereby increase the percentage of oxygen delivered to the tissues.
10. (a) The transfusion would increase the number of red blood cells, which increases the oxygen-carrying capacity of the blood, allowing more sustained effort.
(b) BPG stabilizes the T-state of hemoglobin, which results in a more efficient release of oxygen. If BPG is depleted, the oxygen will not be released even though the red blood cells are carrying more oxygen.
11. Oxygen binding appears to cause the copper ions and their associated histidine ligands to move closer to one another, thereby also moving the helices to which the histidines are attached (in similar fashion to the conformational change in hemoglobin).
12. The modified hemoglobin should not show cooperativity. Although the imidazole in solution will bind to the heme iron (in place of histidine) and will facilitate oxygen binding, the imidazole lacks the crucial connection to the particular α helix that must move so as to transmit the change in conformation.
13. Inositol pentaphosphate (part *c*) is highly anionic, much like 2,3-bisphosphoglycerate.
14.



15. Release of acid will lower the pH. A lower pH promotes oxygen dissociation in the tissues. However, the enhanced release of oxygen in the tissues will increase the concentration of deoxy-Hb, thereby increasing the likelihood that the cells will sickle.
16. (a) $Y = 0.5$ when $pO_2 = 10$ torr. The plot of Y versus pO_2 appears to indicate little or no cooperativity.



(b) The Hill plot shows slight cooperativity with $n = 1.3$ in the central region.



(c) Deoxy dimers of lamprey hemoglobin could have lower affinity for oxygen than do the monomers. If the binding of the first oxygen atom to a dimer causes dissociation of the dimer to give two monomers, then the process would be cooperative. In this mechanism, oxygen binding to each monomer would be easier than binding the first oxygen atom to a deoxy dimer.
17. (a) 2; (b) 4; (c) 2; (d) 1.
18. The electrostatic interactions between BPG and hemoglobin would be weakened by competition with water molecules. The T state would not be stabilized.
19. Fetal hemoglobin contains two α chains and two γ chains. The increase in levels of the γ chain provides an alternative to the mutated β chain in sickle-cell anemia patients. Hence, by increasing fetal hemoglobin expression, hydroxyurea reduces the potential for the formation of insoluble hemoglobin aggregates.

Chapter 8

1. Rate enhancement and substrate specificity.
2. A cofactor.
3. Coenzymes and metals.

4. Vitamins are converted into coenzymes.
5. Enzymes facilitate the formation of the transition state.
6. The intricate three-dimensional structure of proteins allows the construction of active sites that will recognize only specific substrates.
7. The energy required to reach the transition state (the activation energy) is returned when the transition state proceeds to product.
8. (a) $K = \frac{[\mathrm{P}]}{[\mathrm{S}]} = \frac{k_\mathrm{F}}{k_\mathrm{R}} = \frac{10^{-4}}{10^{-6}} = 100$. Using equation 5 in the text, $\Delta G^{\circ\prime} = -11.42$ kJ mol^{-1} (-2.73 kcal mol^{-1})

(b) $k_\mathrm{F} = 10^{-2}$ s^{-1} and $k_\mathrm{R} = 10^{-4}$ s^{-1}. The equilibrium constant and $\Delta G^{\circ\prime}$ values are the same for both the uncatalyzed and catalyzed reactions.
9. Protein hydrolysis has a large activation energy. Protein synthesis must require energy to proceed.
10. The enzymes help protect the fluid that surrounds eyes from bacterial infection.
11. Binding energy is the free energy released when two molecules bind together, such as when an enzyme and a substrate interact.
12. Binding energy is maximized when an enzyme interacts with the transition state, thereby facilitating the formation of the transition state and enhancing the rate of the reaction.
13. There would be no catalytic activity. If the enzyme–substrate complex is more stable than the enzyme–transition-state complex, the transition state would not form and catalysis would not take place.
14. Transition states are very unstable. Consequently, molecules that resemble transition states are themselves likely to be unstable and, hence, difficult to synthesize.
15. (a) 0; (b) +28.53; (c) −22.84; (d) −11.42; (e) +5.69.
16. (a) $\Delta G^{\circ\prime} = -RT \ln K'_\mathrm{eq}$

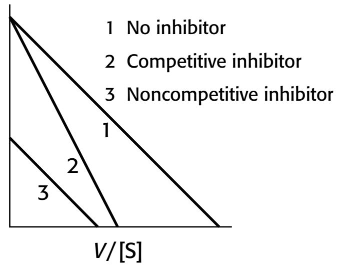
$$+1.8 = -(1.98 \times 10^{-3}\ \mathrm{kcal}^{-1}\ \mathrm{K}^{-1}\ \mathrm{mol}^{-1})(298\ \mathrm{K})\ (\ln[\mathrm{G1P}]/[\mathrm{G6P}])$$

$$-3.05 = \ln[\mathrm{G1P}]/[\mathrm{G6P}]$$

$$+3.05 = \ln[\mathrm{G6P}]/[\mathrm{G1P}]$$

$$K'^{-1}_\mathrm{eq} = 21 \quad \text{or} \quad K'_\mathrm{eq} = 4.8 \times 10^{-2}$$

Because [G6P]/[G1P] = 21, there is 1 molecule of G1P for every 21 molecules of G6P. Because we started with 0.1 M, the [G1P] is 1/22(0.1 M) = 0.0045 M and [G6P] must be 21/22(0.1 M) or 0.096 M. Consequently, the reaction does not proceed as written to a significant extent.
(b) Supply G6P at a high rate and remove G1P at a high rate by other reactions. In other words, make sure that the [G6P]/[G1P] is kept large.
17. $K_\mathrm{eq} = 19$, $\Delta G^{\circ\prime} = -7.3$ kJ mol^{-1} (-1.77 kcal mol^{-1})
18. The three-dimensional structure of an enzyme is stabilized by interactions with the substrate, reaction intermediates, and products. This stabilization minimizes thermal denaturation.
19. At substrate concentrations near the K_M, the enzyme displays significant catalysis yet is sensitive to changes in substrate concentration.
20. No, K_M is not equal to the dissociation constant, because the numerator also contains k_2, the rate constant for the conversion of the enzyme–substrate complex into enzyme and product. If, however, k_2 is much smaller than k_{-1}, $K_\mathrm{M} \approx K_\mathrm{d}$.
21. When [S] = 10 K_M, $V_0 = 0.91\ V_\mathrm{max}$. When [S] = 20 K_M, $V_0 = 95\ V_\mathrm{max}$.
So any Michaelis–Menten curves showing that the enzyme actually attains V_max are pernicious lies.
22. (a) 31.1 μmol; (b) 0.05 μmol; (c) 622 s^{-1}, a midrange value for enzymes (Table 8.5).
23. (a) Yes, $K_\mathrm{M} = 5.2 \times 10^{-6}$ M; (b) $V_\mathrm{max} = 6.8 \times 10^{-10}$ mol minute^{-1}; (c) 337 s^{-1}.
24. Penicillinase, like glycopeptide transpeptidase, forms an acyl-enzyme intermediate with its substrate but transfers the intermediate to water rather than to the terminal glycine residue of the pentaglycine bridge.
25. (a) V_max is 9.5 μmol minute^{-1}. K_M is 1.1×10^{-5} M, the same as without inhibitor.
(b) Noncompetitive.
(c) 2.5×10^{-5} M
(d) $f_\mathrm{ES} = 0.73$, in the presence or absence of this noncompetitive inhibitor.
26. (a) $V_0 = V_\mathrm{max} - (V_0/[\mathrm{S}])\,K_\mathrm{M}$
(b) Slope $= -K_\mathrm{M}$, y-intercept $= V_\mathrm{max}$, x-intercept $= V_\mathrm{max}/K_\mathrm{M}$
(c) An Eadie–Hofstee plot.



27. Sequential reactions are characterized by the formation of a ternary complex consisting of the enzyme and both substrates. Double-displacement reactions always require the formation of a temporarily substituted enzyme intermediate.
28. The rates of utilization of substrates A and B are given by

$$V_\mathrm{A} = \left(\frac{k_\mathrm{cat}}{K_\mathrm{M}}\right)_\mathrm{A}[\mathrm{E}][\mathrm{A}]$$

and

$$V_\mathrm{B} = \left(\frac{k_\mathrm{cat}}{K_\mathrm{M}}\right)_\mathrm{B}[\mathrm{E}][\mathrm{B}]$$

Hence, the ratio of these rates is

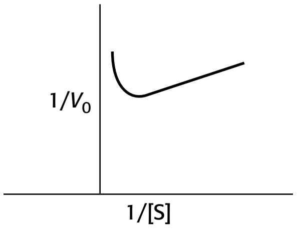
$$V_\mathrm{A}/V_\mathrm{B} = \left(\frac{k_\mathrm{cat}}{K_\mathrm{M}}\right)_\mathrm{B}[\mathrm{A}]/\left(\frac{k_\mathrm{cat}}{K_\mathrm{M}}\right)_\mathrm{A}[\mathrm{B}]$$

Thus, an enzyme discriminates between competing substrates on the basis of their values of $k_\mathrm{cat}/K_\mathrm{M}$ rather than of K_M alone.
29. The mutation slows the reaction by a factor of 100 because the activation free energy is increased by +11.42 kJ mol^{-1} (+2.73 kcal mol^{-1}). Strong binding of the substrate relative to the transition state slows catalysis.
30. 11 μmol minute^{-1}
31. (a) This piece of information is necessary for determining the correct dosage of succinylcholine to administer.
(b) The duration of the paralysis depends on the ability of the serum cholinesterase to clear the drug. If there were one-eighth the amount of enzyme activity, paralysis could last eight times as long.
(c) K_M is the concentration needed by the enzyme to reach ½ V_max. Consequently, for a given concentration of substrate, the reaction catalyzed by the enzyme with the lower K_M will have the higher rate. The mutant patient with the higher K_M will clear the drug at a much lower rate.
32. (a) K_M is a measure of affinity *only* if k_2 is rate limiting, which is the case here. Therefore, the lower K_M means higher affinity. The mutant enzyme has a higher affinity.

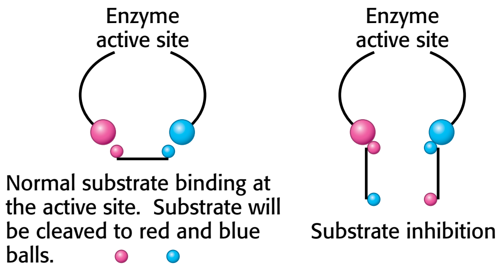
(b) 50 μmol minute^{-1}. 10 mM is K_M, and K_M yields ½ V_{max}. V_{max} is 100 μmol minute^{-1}, and so. . .
(c) Enzymes do not alter the equilibrium of the reaction.
33. Enzyme 2. Despite the fact that enzyme 1 has a higher V_{max} than enzyme 2, enzyme 2 shows greater activity at the concentration of the substrate in the environment because enzyme 2 has a lower K_M for the substrate.
34. (a) The most effective means of measuring the efficiency of any enzyme–substrate complex is to determine the k_{cat}/K_M values. For the three substrates in question, the respective values of k_{cat}/K_M are: 6, 15, and 36. Thus, the enzyme exhibits a strong preference for cleaving peptide bonds in which the second amino acid is a large hydrophobic amino acid.
(b) The k_{cat}/K_M for this substrate is 2. Not very effective. This value suggests that the enzyme prefers to cleave peptide bonds with the following specificity: small R group—large hydrophobic R group.
35. If the total amount of enzyme (E_T) is increased, V_{max} will increase, because $V_{max} = k_2[E]_T$. But $K_M = (k_{-1} + k_2)/k_1$; that is, it is independent of substrate concentration. The middle graph describes this situation.
36.

Experimental condition	V_{max}	K_M
a. Twice as much enzyme is used	Doubles	No change
b. Half as much enzyme is used	Half as large	No change
c. A competitive inhibitor is present	No change	Increases
d. An uncompetitive inhibitor is present	Decreases	Decreases
e. A pure non-competitive is present	Decreases	No change

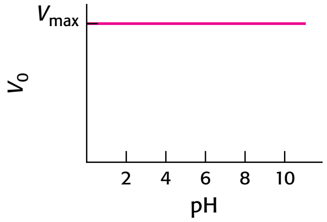
37. (a)

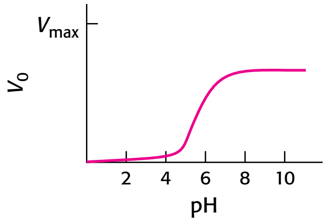


(b) This behavior is substrate inhibition: at high concentrations, the substrate forms unproductive complexes at the active site. The adjoining drawing shows what might happen. Substrate normally binds in a defined orientation, shown in the drawing as red to red and blue to blue. At high concentrations, the substrate may bind at the active site such that the proper orientation is met for each end of the molecule, but two different substrate molecules are binding.


Enzyme active site
Normal substrate binding at the active site. Substrate will be cleaved to red and blue balls.
Enzyme active site
Substrate inhibition

38. The first step will be the rate-limiting step. Enzymes E_B and E_C are operating at ½ V_{max}, whereas the K_M for enzyme E_A is greater than the substrate concentration. E_A would be operating at approximately $10^{-2}V_{max}$.
39. The fluorescence spectroscopy reveals the existence of an enzyme–serine complex and of an enzyme–serine–indole complex.
40. (a) When [S^+] is much greater than the value of K_M, pH will have a negligible effect on the enzyme because S^+ will interact with E^- as soon as the enzyme becomes available.





(b) When [S^+] is much less than the value of K_M, the plot of V_0 versus pH becomes essentially a titration curve for the ionizable groups, with enzyme activity being the titration marker. At low pH, the high concentration of H^+ will keep the enzyme in the EH form and inactive. As the pH rises, more and more of the enzyme will be in the E^- form and active. At high pH (low H^+), all of the enzyme is E^-.
(c) The midpoint on this curve will be the pK_a of the ionizable group, which is stated to be pH 6.
41. (a) Incubating the enzyme at 37°C leads to a denaturation of enzyme structure and a loss of activity. For this reason, most enzymes must be kept cool if they are not actively catalyzing their reactions.
(b) The coenzyme apparently helps to stabilize enzyme structure, because enzyme from PLP-deficient cells denatures faster. Cofactors often help stabilize enzyme structure.
42. The slow second order step occurs when complementary sequences on two different molecules bind with one another. Once the initial sequence alignment occurs, the remainder of the molecule can quickly locate complementary sequences from this nucleation site and reanneal, a first order reaction.

Chapter 9

1. For the amide substrate, the formation of the acyl-enzyme intermediate is slower than the hydrolysis of the acyl-enzyme intermediate, and so no burst is observed. A burst is observed for ester substrates; the formation of the acyl-enzyme intermediate is faster, leading to the observed burst.
2. The histidine residue in the substrate can substitute to some extent for the missing histidine residue of the catalytic triad of the mutant enzyme.
3. No. The catalytic triad works as a unit. After this unit has been made ineffective by the mutation of histidine to alanine, the further mutation of serine to alanine should have only a small effect.
4. The substitution corresponds to one of the key differences between trypsin and chymotrypsin, and so trypsin-like specificity (cleavage after lysine and arginine) might be predicted.

In fact, additional changes are required to effect this specificity change.

5. Imidazole is apparently small enough to reach the active site of carbonic anhydrase and compensate for the missing histidine. Buffers with large molecular components cannot do so, and the effects of the mutation are more evident.

6. No. The odds of such a sequence being present are approximately 1 in $4^{10} = 1{,}048{,}576$. Because a typical viral genome has only 50,000 bp, the target sequence would be unlikely to be present.

7. No, because the enzyme would destroy the host DNA before protective methylation could take place.

8. No. The bacteria receiving the enzyme would have their own DNA destroyed because they would likely lack the appropriate protective methylase.

9. EDTA will bind to Zn^{2+} and remove the ion, which is required for enzyme activity, from the enzyme.

10. (a) The aldehyde reacts with the active-site serine. (b) A hemiacetal is formed.

11. Trypsin.

12. The reaction is expected to be slower by a factor of 10 because the rate depends on the pK_a of the zinc-bound water. $k_{cat} = 60{,}000\ s^{-1}$.

13. EDTA binds the magnesium necessary for the reaction.

14. ATP hydrolysis is reversible within the active site. ATP hydrolysis takes place within the active site with the incorporation of ^{18}O, ATP is re-formed, and the ATP is released back into solution.

15. If the aspartate is mutated, the protease is inactive and the virus will not be viable.

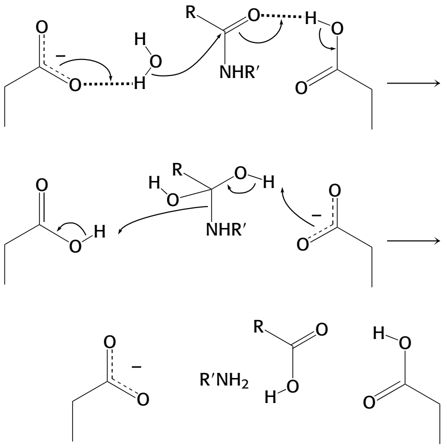
16. Water substitutes for the hydroxyl group of serine 236 in mediating proton transfer from the attacking water and the γ-phosphoryl group.

17. For subtilisin, the catalytic power is approximately $30\ s^{-1}/10^{-8}\ s^{-1} = 3 \times 10^9$. For carbonic anhydrase (at pH 7), the catalytic power is approximately $500{,}000\ s^{-1}/0.15\ s^{-1} = 3.3 \times 10^6$. Subtilisin is the more powerful enzyme by this criterion.

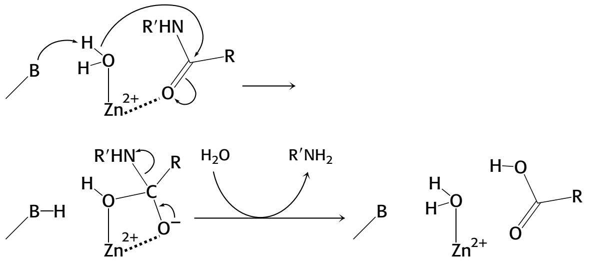
18. The triple mutant still catalyzes the reaction by a factor of approximately 1000-fold compared to the uncatalyzed reaction. This could be due to the enzyme binding the substrate and holding it in a conformation that is susceptible to attack by water.

19. (a) Cysteine protease: The same as Figure 9.8, except that cysteine replaces serine in the active site and no aspartate is present.

(b) Aspartyl protease:



(c) Metalloprotease:



Chapter 10

1. The enzyme catalyzes the first step in the synthesis of pyrimidines. It facilitates the condensation of carbamoyl phosphate and aspartate to form *N*-carbamoylaspartate and inorganic phosphate.

2. The protonated form of histidine probably stabilizes the negatively charged carbonyl oxygen atom of the scissile bond (bond to be broken) in the transition state. Deprotonation would lead to a loss of activity. Hence, the rate is expected to be half maximal at a pH of about 6.5 (the p*K* of an unperturbed histidine side chain in a protein) and to decrease as the pH is raised.

3. The inhibition of an allosteric enzyme by the end product of the pathway controlled by the enzyme. It prevents the production of too much end product and the consumption of substrates when product is not required.

4. High concentrations of ATP might signal two overlapping situations. The high levels of ATP might suggest that some *nucleotides* are available for nucleic acid synthesis, and consequently, UTP and CTP should be synthesized. The high levels of ATP indicate that *energy* is available for nucleic acid synthesis, and so UTP and CTP should be produced.

5. All of the enzyme would be in the R form all of the time. There would be no cooperativity. The kinetics would look like that of a Michaelis–Menten enzyme.

6. The enzyme would show simple Michaelis–Menten kinetics because it is essentially always in the R state.

7. CTP is formed by the addition of an amino group to UTP. Evidence indicates the UTP is also capable of inhibiting ATCase in the presence of CTP.

8. Homotropic effectors are the substrates of allosteric enzymes. Heterotropic effectors are the regulators of allosteric enzymes. Homotropic effectors account for the sigmoidal nature of the velocity versus substrate concentration curve, whereas heterotropic effectors alter the K_M of the curve. Ultimately, both types of effectors work by altering the T/R ratio.

9. The reconstitution shows that the complex quaternary structure and the resulting catalytic and regulatory properties are ultimately encoded in the primary structure of individual components.

10. If substrates had been used, the enzyme would catalyze the reaction. Intermediates would not accumulate on the enzyme. Consequently, any enzyme that crystallized would have been free of substrates or products.

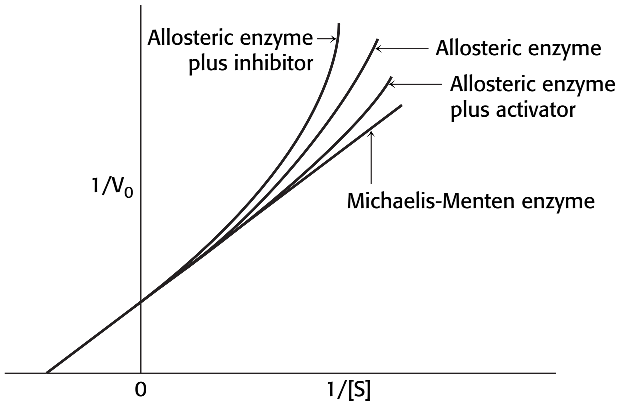
11. (a) 100. The change in the [R]/[T] ratio on binding one substrate molecule must be the same as the ratio of the substrate affinities of the two forms.

(b) 10. The binding of four substrate molecules changes the [R]/[T] by a factor of $100^4 = 10^8$. The ratio in the absence of substrate is 10^{-7}. Hence, the ratio in the fully liganded molecule is $10^8 \times 10^{-7} = 10$.

12. The fraction of molecules in the R form is 10^{-5}, 0.004, 0.615, 0.998, and 1 when 0, 1, 2, 3, and 4 ligands, respectively, are bound.

13. The sequential model can account for negative cooperativity, whereas the concerted model cannot.

14.



15. The binding of PALA switches ATCase from the T to the R state because PALA acts as a substrate analog. An enzyme molecule containing bound PALA has fewer free catalytic sites than does an unoccupied enzyme molecule. However, the PALA-containing enzyme will be in the R state and, hence, have a higher affinity for the substrates. The dependence of the degree of activation on the concentration of PALA is a complex function of the allosteric constant L_0 and of the binding affinities of the R and T states for the analog and substrates.

16. The net outcome of the two reactions is the hydrolysis of ATP to ADP and P_i, which has a ΔG of -50 kJ mol^{-1} (-12 kcal mol^{-1}) under cellular conditions.

17. Isozymes are homologous enzymes that catalyze the same reaction but have different kinetic or regulatory properties.

18. Although the same reaction may be required in a variety of tissues, the biochemical properties of tissues will differ according to their biological function. Isozymes allow the fine-tuning of catalytic and regulatory properties to meet the specific needs of the tissue.

19. (a) 7; (b) 8; (c) 11; (d) 6; (e) 1; (f) 12; (g) 3; (h) 4; (i) 5; (j) 2; (k) 10; (l) 9.

20. When phosphorylation takes place at the expense of ATP, sufficient energy is expended to dramatically alter the structure and hence activity of a protein. Moreover, because ATP is the cellular energy currency, protein modification is linked to the energy status of the cell.

21. Covalent modification is reversible, whereas proteolytic cleavage is irreversible.

22. Activation is independent of zymogen concentration because the reaction is intramolecular.

23. Although quite rare, cases of enteropeptidase deficiency have been reported. The affected person has diarrhea and fails to thrive because digestion is inadequate. In particular, protein digestion is impaired.

24. Add blood from the second patient to a sample from the first. If the mixture clots, the second patient has a defect different from that of the first. This type of assay is called a complementation test.

25. Activated factor X remains bound to blood-platelet membranes, which accelerates its activation of prothrombin.

26. Antithrombin III is a very slowly hydrolyzed substrate of thrombin. Hence, its interaction with thrombin requires a fully formed active site on the enzyme.

27. Replacing methionine with leucine would be a good choice. Leucine is resistant to oxidation and has nearly the same volume and degree of hydrophobicity as methionine has.

28. Inappropriate clot formation could block arteries in the brain, causing a stroke, or the heart, causing a heart attack.

29. Thrombin catalyzes the hydrolysis of fibrinogen to form active fibrin. But it also has a role in shutting down the cascade by activating protein C, a protease that digests other clotting enzymes V_a and $VIII_a$.

30. Tissue-type plasminogen activator, or TPA, is a serine protease that leads to the dissolution of blood clots. TPA activates plasminogen that is bound to a fibrin clot, converting it into active plasmin, which then hydrolyzes the fibrin of the clot.

31. A mature clot is stabilized by amide linkages between the side chains of lysine and glutamine that are absent in a soft clot. The linkages are formed by transglutaminase.

32. The simple sequential model predicts that the fraction of catalytic chains in the R state, f_R, is equal to the fraction containing bound substrate, Y. The concerted model, in contrast, predicts that f_R increases more rapidly than Y as the substrate concentration is increased. The change in f_R leads to the change in Y on addition of substrate, as predicted by the concerted model.

33. The binding of succinate to the functional catalytic sites of the native c_3 moiety changed the visible absorption spectrum of nitrotyrosine residues in the *other* c_3 moiety of the hybrid enzyme. Thus, the binding of substrate analog to the active sites of one trimer altered the structure of the other trimer.

34. According to the concerted model, an allosteric activator shifts the conformational equilibrium of all subunits toward the R state, whereas an allosteric inhibitor shifts it toward the T state. Thus, ATP (an allosteric activator) shifted the equilibrium to the R form, resulting in an absorption change similar to that obtained when substrate is bound. CTP had a different effect. Hence, this allosteric inhibitor shifted the equilibrium to the T form. Thus, the concerted model accounts for the ATP-induced and CTP-induced (heterotropic), as well as for the substrate-induced (homotropic), allosteric interactions of ATCase.

35. (a) When crowded, the control group displays gregarious behavior. (b) Inhibition of PKA appears to prevent gregarious behavior, while inhibition of PKG has no effect on the behavior. (c) The effect of PKG inhibition was investigated to establish that the effect seen with PKA inhibition is specific and not just due to the inhibition of any kinase. (d) PKA plays a role in altering behavior of insects. (e) Gregariousness in locusts that are always gregarious is not affected by PKA inhibition, suggesting that PKA may play a role in modifying behavior patterns but not in establishing the patterns.

36. Residues *a* and *d* are located in the interior of an α-helical coiled coil, near the axis of the superhelix. Hydrophobic interactions between these side chains contribute to the stability of the coiled coil.

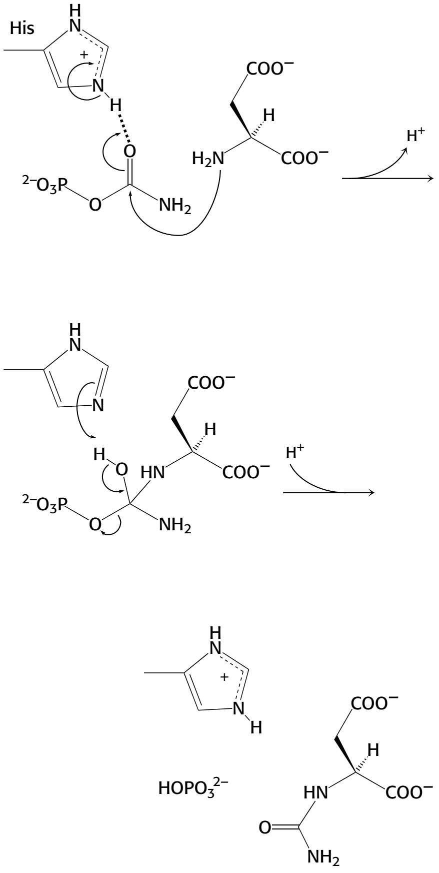
37. In the R state, ATCase expands and becomes less dense. This decrease in density results in a decrease in the sedimentation value (see the formula on p. 76).

38. The interaction between trypsin and the inhibitor is so stable that the transition state is rarely formed. Recall that maximal binding energy is released when an enzyme binds to the transition state. If the substrate-enzyme interaction is too stable, the transition state rarely forms.

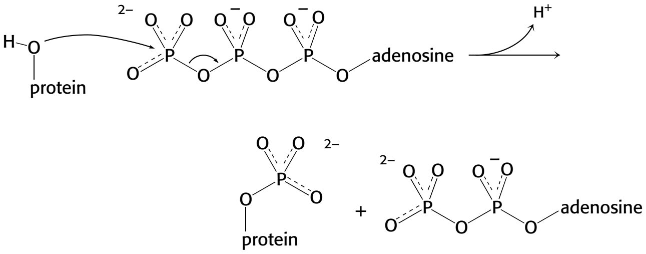
39. Dicoumarol is a competitive inhibitor of γ-glutamyl carboxylase. Consequently, γ-carboxyglutamate, which is required for prothrombin to be converted into thrombin, is not formed. Amino acid composition is determined by

hydrolyzing the protein at elevated temperatures in the presence of strong acid. Under these conditions, the carboxyl group of γ-carboxyglutamate is removed from the normal prothrombin, leaving simply glutamate.

40.


His
H
N
+
N
H
O
$^{2-}O_3P$
O
NH_2
COO^-
H
H_2N
COO^-
H^+
H
N
N
COO^-
H
H
O
HN
COO^-
H^+
$^{2-}O_3P$
O
NH_2
H
N
+
N
H
COO^-
H
$HOPO_3^{2-}$
HN
COO^-
O
NH_2

41.


2–
O O O – O O – O
H
O
protein
O
P
O
P
O
P
O
adenosine
H^+
O O 2–
P
O
O
protein
+
2– O O O – O
O
P
O
P
O
adenosine

Chapter 11

1. Carbohydrates were originally regarded as *hydrates* of *carbon* because the empirical formula of many of them is $(CH_2O)_n$.

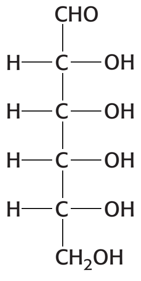
2. Three amino acids can be linked by peptide bonds in only six different ways. However, three different monosaccharides can be linked in a plethora of ways. The monosaccharides can be linked in a linear or branched manner, with α or β linkages, with bonds between C-1 and C-3, between C-1 and C-4, between C-1 and C-6, and so forth. In fact, the three monosaccharides can form 12,288 different trisaccharides.

3. (a) 10; (b) 6; (c) 8; (d) 9; (e) 2; (f) 4; (g) 1; (h) 5; (i) 7; (j) 3.

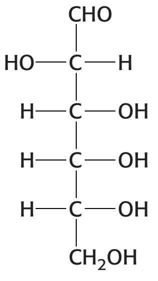
4. (a) aldose-ketose; (b) epimers; (c) aldose-ketose (d) anomers; (e) aldose-ketose; (f) epimers.

5. Erythrose: tetrose aldose; ribose: pentose aldose; glyceraldehyde: triose aldose; dihydroxyacetone: triose ketose; erythrulose: tetrose ketose; ribulose: pentose ketose; fructose: hexose ketose.

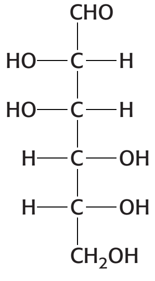
6.



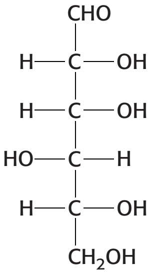
D-Allose



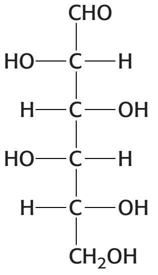
D-Altrose



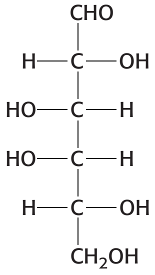
D-Mannose



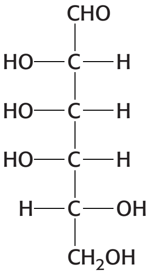
D-Gulose



D-Idose

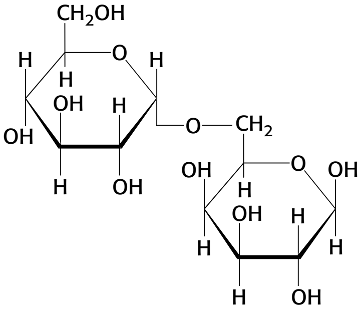


D-Galactose



D-Talose

7.


CH_2OH
H
O
H
H
OH
H
OH
H
OH
O
CH_2
OH
O
OH
H
OH
H
H
H
H
OH

α-Glucosyl-(1 ⟶ 6)-galactose

8. The proportion of the α anomer is 0.36, and that of the β anomer is 0.64.

9. Glucose is reactive because of the presence of an aldehyde group in its open-chain form. The aldehyde group slowly condenses with amino groups to form aldimine products of a type called Schiff-base adducts.

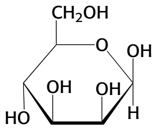
10. A pyranoside reacts with two molecules of periodate; formate is one of the products. A furanoside reacts with only one molecule of periodate; formate is not formed.

11. (a) β-D-Mannose; (b) β-D-galactose; (c) β-D-fructose; (d) β-D-glucosamine.

12. The trisaccharide itself should be a competitive inhibitor of cell adhesion if the trisaccharide unit of the glycoprotein is critical for the interaction.

13. Reducing ends would form 1,2,3,6-tetramethylglucose. The branch points would yield 2,3-dimethylglucose. The remainder of the molecule would yield 2,3,6-trimethylglucose.

14. (a) Not a reducing sugar; no open-chain forms are possible. (b) D-galactose, D-glucose, D-fructose. (c) D-galactose and sucrose (glucose + fructose).



β-D-Mannose

15. The hemiketal linkage of the α anomer is broken to form the open form. Rotation about the C-1 and C-2 bonds allows the formation of the β anomer, and a mixture of isomers results.

16. Heating converts the very sweet pyranose form into the more-stable but less-sweet furanose form. Consequently, the sweetness of the preparation is difficult to accurately control, which also accounts for why honey loses sweetness with time. See Figure 11.5 for structures.

17. (a) Each glycogen molecule has one reducing end, whereas the number of nonreducing ends is determined by the number of branches, or α-1,6 linkages. (b) Because the number of nonreducing ends greatly exceeds the number of reducing ends in a collection of glycogen molecules, all of the degradation and synthesis of glycogen takes place at the nonreducing ends, thus maximizing the rate of degradation and synthesis.

18. No, sucrose is not a reducing sugar. The anomeric carbon atom acts as the reducing agent in both glucose and fructose but, in sucrose, the anomeric carbon atoms of fructose and glucose are joined by a covalent bond and are thus not available to react.

19. Glycogen is a polymer of glucose linked by α-1,4-glycosidic bonds with branches formed approximately every 10 glucose units by α-1,6-glycosidic bonds. Starch consists of two polymers of glucose. Amylose is a straight-chain polymer formed by α-1,4-glycosidic bonds. Amylopectin is similar to glycogen but amylopectin has fewer branches, one branch per 30 or so glucose units.

20. Cellulose is a linear polymer of glucose joined by β-1,4 linkages. Glycogen is a branched polymer with the main chain being formed by α-1,4-glycosidic bonds. The β-1,4 linkages allow the formation of a linear polymer ideal for structural roles. The α-1,4 linkages of glycogen form a helical structure, which allows the storage of many glucose moieties in a small space.

21. Simple glycoproteins are often secreted proteins and thus play a variety of roles. For example, the hormone EPO is a glycoprotein. Usually, the protein component constitutes the bulk of the glycoprotein by mass. In contrast, proteoglycans and mucoproteins are predominantly carbohydrates. Proteoglycans have glycosaminoglycans attached, and play structural roles as in cartilage and the extracellular matrix. Mucoproteins often serve as lubricants and have multiple carbohydrates attached through an *N*-acetylgalactosamine moiety.

22. The attachment of the carbohydrate allows the EPO to stay in circulation longer and thus to function for longer periods of time than would a carbohydrate-free EPO.

23. The glycosaminoglycan, because it is heavily charged, binds many water molecules. When cartilage is stressed, such as when your heel hits the ground, the water is released, thus cushioning the impact. When you lift your heel, the water rebinds.

24. The lectin that binds the mannose 6-phosphate might be defective and not recognize a correctly addressed protein.

25. Asparagine, serine, and threonine.

26. Different molecular forms of a glycoprotein that differ in the amount of carbohydrate attached or the location of attachment or both.

27. The total collection of carbohydrates synthesized by a cell at particular times and under particular environmental conditions.

28. The genome comprises all of the genes present in an organism. The proteome includes all of the possible protein products and modified proteins that a cell expresses under any particular set of circumstances. The glycome consists of all of the carbohydrates synthesized by the cell under any particular set of circumstances. Because the genome is static, but any given protein can be variously expressed and modified, the proteome is more complex than the genome. The glycome, which includes not only glycoforms of proteins, but also many possible carbohydrate structures, must be even more complex.

29. An asparagine residue can be glycosylated if the residue is part of an Asn-X-Ser or Asn-X-Thr sequence, in which X can be any residue, except proline. Your roommate's hedging is because not all potential sites are glycosylated.

30. It suggests that carbohydrates are on the cell surfaces of all organisms for the purpose of recognition by other cells, organisms, or the environment.

31. A glycoprotein is a protein that is decorated with carbohydrates. A lectin is a protein that specifically recognizes carbohydrates. A lectin can also be a glycoprotein.

32. Each site either is or is not glycosylated, and so there are $2^6 = 64$ possible proteins.

33. The 20 amino acid components of proteins and the 4 nucleotide components of nucleic acids are linked by the same type of bond—the peptide bond in proteins and the 5′-to-3′ phosphodiester linkage in nucleic acids. In contrast, there are many different carbohydrates that can be modified and linked in a variety of ways. Moreover, oligosaccharides can be branched. Finally, many sugars have the same or similar chemical formula, and similar chemical properties, making specific identification and linkage difficult.

34. As discussed in Chapter 9, many enzymes display stereochemical specificity. Clearly, the enzymes of sucrose synthesis are able to distinguish between the isomers of the substrates and link only the correct pair.

35. If the carbohydrate specificity of the lectin is known, an affinity column with the appropriate carbohydrate attached could be prepared. The protein preparation containing the lectin of interest could be passed over the column. The use of this method was indeed how the glucose-binding lectin concanavalin A was purified.

36. (a) Aggrecan is heavily decorated with glycosaminoglycans. If glycosaminoglycans are released into the media, aggrecan must be undergoing degradation.

(b) Another enzyme might be present that cleaves glycosaminoglycans from aggrecan without degrading aggrecan. Other experiments not shown established that glycosaminoglycan release is an accurate measure of aggrecan destruction.

(c) The control provides a baseline of "background" degradation inherent in the assay.

(d) Aggrecan degradation is greatly enhanced.

(e) Aggrecan degradation is reduced to the background system.
(f) It is an in vitro system in which not all the factors contributing to cartilage stabilization in vivo are present.

Chapter 12

1. 2.86×10^6 molecules, because each leaflet of the bilayer contains 1.43×10^6 molecules.
2. Essentially an "inside-out" membrane. The hydrophilic groups would come together on the interior of the structure, away from the solvent, whereas the hydrocarbon chains would interact with the solvent.
3. 2×10^{-7} cm, 6×10^{-6} cm, and 2×10^{-4} cm.
4. The radius of this molecule is 3.1×10^{-7} cm, and its diffusion coefficient is 7.4×10^{-9} cm^2 s^{-1}. The average distances traversed are 1.7×10^{-7} cm in 1 μs, 5.4×10^{-6} cm in 1 ms, and 1.7×10^{-4} cm in 1 s.
5. The membrane underwent a phase transition from a highly fluid to a nearly frozen state when the temperature was lowered. A carrier can shuttle ions across a membrane only when the bilayer is highly fluid. A channel, in contrast, allows ions to traverse its pore even when the bilayer is quite rigid.
6. The presence of a cis double bond introduces a kink in the fatty acid chain that prevents tight packing and reduces the number of atoms in van der Waals contact. The kink lowers the melting point compared with that of a saturated fatty acid. Trans fatty acids do not have the kink, and so their melting temperatures are higher, more similar to those of saturated fatty acids. Because trans fatty acids have no structural effect, they are rarely observed.
7. Palmitic acid is shorter than stearic acid. Thus, when the chains pack together, there is less opportunity for van der Waals interaction and the melting point is thus lower than that of the longer stearic acid.
8. Hibernators selectively feed on plants that have a high proportion of polyunsaturated fatty acids with lower melting temperature.
9. The initial decrease in fluorescence with the first addition of sodium dithionite results from the quenching of NBD-PS molecules in the outer leaflet of the bilayer. Sodium dithionite does not traverse the membrane under these experimental conditions; hence, it does not quench the labeled phospholipids in the inner leaflet. A second addition of sodium dithionite has no effect, as the NBD-PS molecules in the outer leaflet remain quenched. However, after a 6.5-hour incubation, about half the NBD-PS has flipped over to the outer leaflet of the bilayer, resulting in the 50% decrease in fluorescence when sodium dithionite is added.
10. The addition of the carbohydrate introduces a significant energy barrier to the flip-flop because a hydrophilic carbohydrate moiety would need to be moved through a hydrophobic environment. This energetic barrier enhances membrane asymmetry.
11. The C_{16} alkyl chain is attached by an ether linkage. The C-2 carbon atom of glycerol has only an acetyl group attached by an ester linkage instead of a fatty acid, as is the case with most phospholipids.
12. In a hydrophobic environment, the formation of intrachain hydrogen bonds stabilizes the amide hydrogen atoms and carbonyl oxygen atoms of the polypeptide chain, and so an α helix forms. In an aqueous environment, these groups are stabilized by interaction with water, and so there is no energetic reason to form an α helix. Thus, the α helix would be more likely to form in a hydrophobic environment.
13. The protein may contain an α helix that passes through the hydrophobic core of the protein. This helix is likely to feature a stretch of hydrophobic amino acids similar to those observed in transmembrane helices.
14. The shift to the lower temperature would decrease fluidity by enhancing the packing of the hydrophobic chains by van der Waals interactions. To prevent this packing, new phospholipids having shorter chains and a greater number of cis double bonds would be synthesized. The shorter chains would reduce the number of van der Waals interactions, and the cis double bonds, which cause the kink in structure, would prevent the packing of the fatty acid tails of the phospholipids.
15. Each of the 21 v-SNARE proteins could interact with each of 7 t-SNARE partners. Multiplication gives the total number of different interacting pairs: $7 \times 21 = 147$ different v-SNARE–t-SNARE pairs.
16. (a) The graph shows that, as temperature increases, the phospholipid bilayer becomes more fluid. T_m is the temperature of the transition from the predominantly less fluid state to the predominantly more fluid state. Cholesterol broadens the transition from the less-fluid to the more-fluid state. In essence, cholesterol makes membrane fluidity less sensitive to temperature changes.
(b) This effect is important because the presence of cholesterol tends to stabilize membrane fluidity by preventing sharp transitions. Because protein function depends on the proper fluidity of the membrane, cholesterol maintains the proper environment for membrane-protein function.
17. The protein plotted in part *c* is a transmembrane protein from *C. elegans.* It spans the membrane with four α helices that are prominently displayed as hydrophobic peaks in the hydropathy plot. Interestingly, the protein plotted in part *a* also is a membrane protein, a porin. This protein is made primarily of β strands, which lack the prominent hydrophobic window of membrane helices. This example shows that, although hydropathy plots are useful, they are not infallible.
18. (a) Prostaglandin H_2 synthase-1 recovers its activity immediately after removal of ibuprofen, suggesting that this inhibitor dissociates rapidly from the enzyme. In contrast, the enzyme remains significantly inhibited 30 minutes after removal of indomethacin, suggesting that this inhibitor dissociates slowly from its active site.
(b) Aspirin covalently modifies prostaglandin H_2 synthase-1, indicating that it would dissociate very slowly (if at all). Hence, one would anticipate that very low activity would be evident in all conditions where inhibitor has been added (columns 2, 3, and 4).
19. To purify any protein, the protein must first be solubilized. For a membrane protein, solubilization usually requires a detergent—hydrophobic molecules that bind to the protein and thus replace the lipid environment of the membrane. If the detergent is removed, the protein aggregates and precipitates from solution. Often, the steps in purification, such as ion-exchange chromatography, are difficult to perform in the presence of sufficient detergent to solubilize the protein. Crystals of appropriate protein–detergent complexes must be generated.

Chapter 13

1. In simple diffusion, the substance in question can diffuse down its concentration gradient through the membrane. In facilitated diffusion, the substance is not lipophilic and cannot directly diffuse through the membrane. A channel or carrier is required to facilitate movement down the gradient.

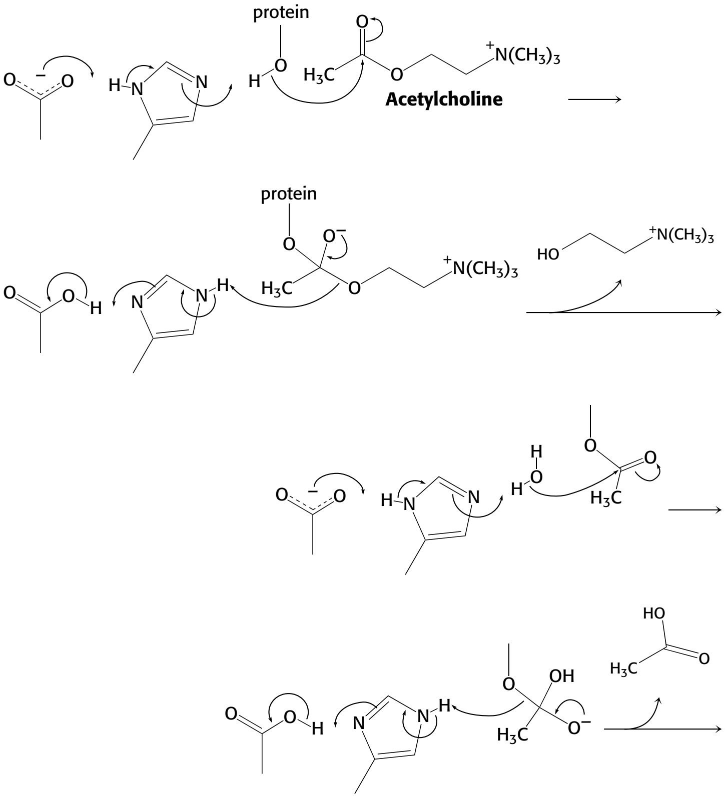
2. The two forms are (1) ATP hydrolysis and (2) the movement of one molecule down its concentration gradient coupled with the movement of another molecule up its concentration gradient.
3. The three types of carriers are symporters, antiporters, and uniporters. Symporters and antiporters can mediate secondary active transport.

4. The free-energy cost is $+32$ kJ mol^{-1} ($+7.6$ kcal mol^{-1}). The chemical work performed is $+20.4$ kJ mol^{-1} ($+4.9$ kcal mol^{-1}), and the electrical work performed is $+11.5$ kJ mol^{-1} ($+2.8$ kcal mol^{-1}).
5. For chloride, $z = -1$; for calcium $z = +2$. At the concentrations given, the equilibrium potential for chloride is -97 mV and the equlibrium potential for calcium is $+122$ mV.
6. The concentration of glucose inside the cell is 66 times as great as that outside the cell $[(c_2/c_1) = 66]$ when the free-energy input is $+10.8$ kJ mol^{-1} ($+2.6$ kcal mol^{-1}).
7. By analogy with the Ca^{2+} ATPase, with three Na^+ ions binding from inside the cell to the E_1 conformation and with two K^+ ions binding from outside the cell to the E_2 conformation, a plausible mechanism is as follows:

(i) A catalytic cycle could begin with the enzyme in its unphosphorylated state (E_1) with three sodium ions bound.
(ii) The E_1 conformation binds ATP. A conformational change traps sodium ions inside the enzyme.
(iii) The phosphoryl group is transferred from ATP to an aspartyl residue.
(iv) On ADP release, the enzyme changes its overall conformation, including the membrane domain. This new conformation (E_2) releases the sodium ions to the side of the membrane opposite that at which they entered and binds two potassium ions from the side where sodium ions are released.
(v) The phosphorylaspartate residue is hydrolyzed to release inorganic phosphate. With the release of phosphate, the interactions stabilizing E_2 are lost, and the enzyme everts to the E_1 conformation. Potassium ions are released to the cytoplasmic side of the membrane. The binding of three sodium ions from the cytoplasmic side of the membrane completes the cycle.

8. Establish a lactose gradient across vesicle membranes that contain properly oriented lactose permease. Initially, the pH should be the same on both sides of the membrane and the lactose concentration should be higher on the "exit" side of lactose permease. As the lactose flows "in reverse" through the permease, down its concentration gradient, it can be tested whether or not a pH gradient becomes established as the lactose gradient is dissipated.
9. Ligand-gated channels open in response to the binding of a molecule by the channel, whereas voltage-gated channels open in response to changes in the membrane potential.
10. An ion channel must transport ions in either direction at the same rate. The net flow of ions is determined only by the composition of the solutions on either side of the membrane.
11. Uniporters act as enzymes do; their transport cycles include large conformational changes, and only a few molecules interact with the protein per transport cycle. In contrast, channels, after having opened, provide a pore in the membrane through which many ions may pass. As such, channels mediate transport at a much higher rate than do uniporters.
12. FCCP effectively creates a pore in the bacterial membrane through which protons can pass rapidly. Protons that are pumped out of the bacteria will pass through this pore preferentially (the "path of least resistance"), rather than participate in H^+/lactose symport.
13. Cardiac muscle must contract in a highly coordinated manner in order to pump blood effectively. Gap junctions mediate the orderly cell-to-cell propagation of the action potential through the heart during each beat.
14. The positively charged guanidinium group resembles Na^+ and binds to negatively charged carboxylate groups in the mouth of the channel.
15. SERCA, a P-type ATPase, uses a mechanism by which a covalent phosphorylated intermediate (at an aspartate residue) is formed. At steady state, a subset of the SERCA molecules is trapped in the E_2-P state and, as a result, radiolabeled. The MDR protein is an ABC transporter and does not operate through a phosphorylated intermediate. Hence, a radiolabeled band would not be observed for MDR.
16. The blockage of ion channels inhibits action potentials, leading to loss of nervous function. Like tetrodotoxin, these toxin molecules are useful for isolating and specifically inhibiting particular ion channels.
17. After repolarization, the ball domains of the ion channels engage the channel pore, rendering them inactive for a short period of time. During this time, the channels cannot be reopened until the ball domains disengage and the channel returns to the "closed" state.
18. Because sodium ions are charged and because sodium channels carry only sodium ions (but not anions), the accumulation of excess positive charge on one side of the membrane dominates the chemical gradients.
19. A mutation that impairs the ability of the sodium channel to inactivate would prolong the duration of the depolarizing sodium current, thus lengthening the cardiac action potential.
20. No. Channels will likely open or close in response to an external stimulus, but the unit conductance of the open channel will be influenced very little.
21. The ratio of closed to open forms of the channel is 10^5, 5000, 250, 12.5, and 0.625 when zero, one, two, three, and four ligands, respectively, are bound. Hence, the fraction of open channels is 1.0×10^{-5}, 2.0×10^{-4}, 4.0×10^{-3}, 7.4×10^{-2}, and 0.62.
22. These organic phosphates inhibit acetylcholinesterase by reacting with the active-site serine residue to form a stable phosphorylated derivative. They cause respiratory paralysis by blocking synaptic transmission at cholinergic synapses.
23. (a) The binding of the first acetylcholine molecule increases the open-to-closed ratio by a factor of 240, and the binding of the second increases it by a factor of 11,700. (b) The free-energy contributions are $+14$ kJ mol^{-1} ($+3.3$ kcal mol^{-1}) and $+23$ kJ mol^{-1} ($+5.6$ kcal mol^{-1}), respectively. (c) No; the MWC model predicts that the binding of each ligand will have the same effect on the open-to-closed ratio.
24. Batrachotoxin blocks the transition from the open to the closed state.
25. (a) Chloride ions flow into the cell. (b) Chloride flux is inhibitory because it hyperpolarizes the membrane. (c) The channel consists of five subunits.
26. After the addition of ATP and calcium, SERCA will pump Ca^{2+} ions into the vesicle. However, the accumulation of Ca^{2+} ions inside the vesicle will rapidly lead to the formation of an electrical gradient that cannot be overcome by ATP hydrolysis. The addition of calcimycin will allow the pumped Ca^{2+} ions to flow back out of the vesicle, dissipating the charge buildup, and enabling the pump to operate continuously.
27. The catalytic prowess of acetylcholinesterase ensures that the duration of the nerve stimulus will be short.

28. See reaction below.



29. (a) Only ASIC1a is inhibited by the toxin. (b) Yes; when the toxin was removed, the activity of the acid-sensing channel began to be restored. (c) 0.9 nM.

30. This mutation is one of a class of mutations that result in slow-channel syndrome (SCS). The results suggest a defect in channel closing; so the channel remains open for prolonged periods. Alternatively, the channel may have a higher affinity for acetylcholine than does the control channel.

31. The mutation reduces the affinity of acetylcholine for the receptor. The recordings would show the channel opening only infrequently.

32. Glucose displays a transport curve that suggests the participation of a carrier because the initial rate is high but then levels off at higher concentrations, consistent with saturation of the carrier, which is reminiscent of Michaelis–Menten enzymes (Section 8.4). Indole shows no such saturation phenomenon, which implies that the molecule is lipophilic and simply diffuses across the membrane. Ouabain is a specific inhibitor the Na^+–K^+ pump. If ouabain were to inhibit glucose transport, then a Na^+-glucose cotransporter would be assisting in transport.

Chapter 14

1. The negatively charged glutamate residues mimic the negatively charged phosphoserine or phosphothreonine residues and stabilize the active conformation of the enzyme.

2. No. Phosphoserine and phosphothreonine are considerably shorter than phosphotyrosine.

3. The GTPase activity terminates the signal. Without such activity, after a pathway has been activated, it remains activated and is unresponsive to changes in the initial signal. If the GTPase activity were more efficient, the lifetime of the GTP-bound G_α subunit would be too short to achieve downstream signaling.

4. Two identical receptor molecules must recognize different aspects of the same signal molecule.

5. Growth-factor receptors can be activated by dimerization. If an antibody causes a receptor to dimerize, the signal-transduction pathway in a cell will be activated.

6. The mutated α subunit will always be in the GTP form and, hence, in the active form, which would stimulate its signaling pathway.

7. A G protein is a component of the signal-transduction pathway. GTPγS is not hydrolyzed by the G_α subunit, leading to prolonged activation.

8. Calcium ions diffuse slowly because they bind to many protein surfaces within a cell, impeding their free motion. Cyclic AMP does not bind as frequently, and so it diffuses more rapidly.

9. Fura-2 is a highly negatively charged molecule, with five carboxylate groups. Its charge prevents it from effectively crossing the hydrophobic region of the plasma membrane.

10. $G_{\alpha s}$ stimulates adenylate cyclase, leading to the generation of cAMP. This signal then leads to glucose mobilization (Chapter 21). If cAMP phosphodiesterase were inhibited, then cAMP levels would remain high even after the termination of the epinephrine signal, and glucose mobilization would continue.

11. If the two kinase domains are forced to be within close proximity of each other, the activation loop of one kinase, in its inactivating conformation, can be displaced by the activation loop of the neighboring kinase, which acts as a substrate for phosphorylation.

12. The full network of pathways initiated by insulin includes a large number of proteins and is substantially more elaborate than indicated in Figure 14.26. Furthermore, many additional proteins take part in the termination of insulin signaling. A defect in any of the proteins in the insulin signaling pathways or in the subsequent termination of the insulin response could potentially cause problems. Therefore, it is not surprising that many different gene defects can cause type 2 diabetes.

13. The binding of growth hormone causes its monomeric receptor to dimerize. The dimeric receptor can then activate a separate tyrosine kinase to which the receptor binds. The signaling pathway can then continue in similar fashion to the pathways that are activated by the insulin receptor or other mammalian EGF receptors.

14. The truncated receptor will dimerize with the full-length monomers on EGF-binding, but cross-phosphorylation cannot take place, because the truncated receptor possesses neither the substrate for the neighboring kinase domain nor its own kinase domain to phosphorylate the C-terminal tail of the other monomer. Hence, these mutant receptors will block normal EGF signaling.

15. Insulin would elicit the response that is normally caused by EGF. Insulin binding will likely stimulate dimerization and phosphorylation of the chimeric receptor and thereby signal the downstream events that are normally triggered by EGF binding. Exposure of these cells to EGF would have no effect.

16. 10^5

17. The formation of diacylglycerol implies the participation of phospholipase C. A simple pathway would entail receptor activation by cross-phosphorylation, followed by the binding of phospholipase C (through its SH2 domains). The participation of phospholipase C indicates that IP_3 would be formed and, hence, calcium concentrations would increase.

18. Other potential drug targets within the EGF signaling cascade include, but are not limited to, the kinase active sites of the EGF receptor, Raf, MEK, or ERK.

19. In the reaction catalyzed by adenylate cyclase, the 3′-OH group nucleophilically attacks the α-phosphorus atom attached to the 5′-OH group, leading to displacement of pyrophosphate. The reaction catalyzed by DNA polymerase is similar except that the 3′-OH group is on a different nucleotide.

20. ATP-competitive inhibitors are likely to act on multiple kinases because every kinase domain contains an ATP-binding site. Hence, these drugs may not be selective for the desired kinase target.

21. (a) $X \approx 10^{-7}$ M; $Y \approx 5 \times 10^{-6}$ M; $Z \approx 10^{-3}$ M. (b) Because much less X is required to fill half of the sites, X displays the highest affinity. (c) The binding affinity almost perfectly matches the ability to stimulate adenylate cyclase, suggesting that the hormone–receptor complex leads to the stimulation of adenylate cyclase. (d) Try performing the experiment in the presence of antibodies to $G_{\alpha s}$.

22. (a) The total binding does not distinguish binding to a specific receptor from binding to different receptors or from nonspecific binding to the membrane.

(b) The rationale is that the receptor will have a high affinity for the ligand. Thus, in the presence of excess nonradioactive ligand, the receptor will bind to nonradioactive ligand. Therefore, any binding of the radioactive ligand must be nonspecific.

(c) The plateau suggests that the number of receptor-binding sites in the cell membrane is limited.

23. Number of receptors per cell =

$$\frac{10^4\ \text{cpm}}{\text{mg of membrane protein}} \times \frac{\text{mg of membrane protein}}{10^{10}\ \text{cells}} \times \frac{\text{mmol}}{10^{12}\ \text{cpm}} \times \frac{6.023 \times 10^{20}\ \text{molecules}}{\text{mmol}} = 600$$

Chapter 15

1. The highly integrated biochemical reactions that take place inside the cell.

2. Anabolism is the set of biochemical reactions that use energy to build new molecules and ultimately new cells. Catabolism is the set of biochemical reactions that extract energy from fuel sources or break down biomolecules.

3. You reply that vandalism is disrespectful and expensive. Part of your tuition money will now have to pay to remove the vandalism. Plus, the fool should know that Gibbs free energy is at a minimum when a system is in equilibrium.

4. Cellular movements and the performance of mechanical work; active transport; biosynthetic reactions.

5. 1. f; 2. h; 3. i; 4. a; 5. g; 6. b; 7. c; 8. e; 9. j; 10. d.

6. These ions neutralize the charges on the ATP and also facilitate interactions with macromolecules that bind ATP.

7. Charge repulsion, resonance stabilization, increase in entropy, and stabilization by hydration.

8. Trick question. The answer is not known. Adenine appears to form more readily under prebiotic conditions, so ATP may have predominated initially.

9. Having only one nucleotide represent the available energy allows the cell to better monitor its energy status.

10. Increasing the concentration of ATP or decreasing the concentration cellular ADP or P_i (by rapid removal by other reactions, for instance) would make the reaction more exergonic. Likewise, altering the Mg^{2+} concentration could raise or lower the ΔG of the reaction.

11. The free-energy changes of the individual steps in a pathway are summed to determine the overall free-energy change of the entire pathway. Consequently, a reaction with a positive free-energy value can be powered to take place if coupled to a sufficiently exergonic reaction.

12. Reactions in parts *a* and *c,* to the left; reactions in parts *b* and *d,* to the right.

13. None whatsoever.

14. (a) $\Delta G^{\circ\prime} = +31.4$ kJ mol^{-1} ($+7.5$ kcal mol^{-1}) and $K'_{\text{eq}} = 3.06 \times 10^{-6}$; (b) 3.28×10^4.

15. $\Delta G^{\circ\prime} = +7.1$ kJ mol^{-1} ($+1.7$ kcal mol^{-1}). The equilibrium ratio is 17.5.

16. (a) Acetate + CoA + H^+ goes to acetyl CoA + H_2O, $\Delta G^{\circ\prime} = -31.4$ kJ mol^{-1} (-7.5 kcal mol^{-1}). ATP hydrolysis to AMP and PP_i, $\Delta G^{\circ\prime} = -45.6$ kJ mol^{-1} (-10.9 kcal mol^{-1}). Overall reaction, $\Delta G^{\circ\prime} = -14.2$ kJ mol^{-1} (-3.4 kcal mol^{-1}).

(b) With pyrophosphate hydrolysis, $\Delta G^{\circ\prime} = -33.4$ kJ mol^{-1} (-7.98 kcal mol^{-1}). Pyrophosphate hydrolysis makes the overall reaction even more exergonic.

17. (a) For an acid AH,

$$\text{AH} \rightleftharpoons \text{A}^- + \text{H}^+ \quad K = \frac{[\text{A}^-][\text{H}^+]}{[\text{AH}]}$$

The pK is defined as p$K = -\log_{10} K$. $\Delta G^{\circ\prime}$ is the standard free-energy change at pH 7. Thus, $\Delta G^{\circ\prime} = -RT \ln K = -2.303\,RT \log_{10} K = +2.303\,RT\,\text{p}K$.

(b) $\Delta G^{\circ\prime} = +27.32$ kJ mol^{-1} ($+6.53$ kcal mol^{-1}).

18. Arginine phosphate in invertebrate muscle, like creatine phosphate in vertebrate muscle, serves as a reservoir of high-potential phosphoryl groups. Arginine phosphate maintains a high level of ATP in muscular exertion.

19. An ADP unit.

20. (a) The rationale behind creatine supplementation is that it would be converted into creatine phosphate and thus serve as a rapid means of replenishing ATP after muscle contraction. (b) If creatine supplementation is beneficial, it would affect activities that depend on short bursts of activity; any sustained activity would require ATP generation by fuel metabolism, which, as Figure 15.7 shows, requires more time.

21. Under standard conditions, $\Delta G^{\circ\prime} = -RT \ln$ [products]/[reactants]. Substituting $+23.8$ kJ mol^{-1} ($+5.7$ kcal mol^{-1}) for $\Delta G^{\circ\prime}$ and solving for [products]/[reactants] yields 9.9×10^{-5}. In other words, the forward reaction does not take place to a significant extent. Under intracellular conditions, ΔG is -1.3 kJ mol^{-1} (-0.3 kcal mol^{-1}). Using the equation $\Delta G = \Delta G^{\circ\prime} + RT \ln$ [products]/[reactants] and solving for [products]/[reactants] gives a ratio of 5.96×10^{-5}. Thus, a reaction that is endergonic under standard conditions can be converted into an exergonic reaction by maintaining the [products]/[reactants] ratio below the equilibrium value. This conversion is usually attained by using the products in another coupled reaction as soon as they are formed.

22. Under standard conditions,

$$K'_{\text{eq}} = \frac{[\text{B}]_{\text{eq}}}{[\text{A}]_{\text{eq}}} \times \frac{[\text{ADP}]_{\text{eq}}[\text{P}_\text{i}]_{\text{eq}}}{[\text{ATP}]_{\text{eq}}} = 10^{33/1.36} = 2.67 \times 10^2$$

At equilibrium, the ratio of [B] to [A] is given by

$$\frac{[\text{B}]_{\text{eq}}}{[\text{A}]_{\text{eq}}} = K'_{\text{eq}} \frac{[\text{ATP}]_{\text{eq}}}{[\text{ADP}]_{\text{eq}}[\text{P}_\text{i}]_{\text{eq}}}$$

The ATP-generating system of cells maintains the [ATP]/[ADP] [P_i] ratio at a high level, typically about 500 M^{-1}. For this ratio,

$$\frac{[\text{B}]_{\text{eq}}}{[\text{A}]_{\text{eq}}} = 2.67 \times 10^2 \times 500 = 1.34 \times 10^5$$

This equilibrium ratio is strikingly different from the value of 1.15×10^{-3} for the reaction $A \rightarrow B$ in the absence of ATP hydrolysis. In other words, coupling the hydrolysis of ATP with the conversion of A into B has changed the equilibrium ratio of B to A by a factor of about 10^8.

23. Liver: -45.2 kJ mol^{-1} (-10.8 kcal mol^{-1}); muscle: -48.1 kJ mol^{-1} (-11.5 kcal mol^{-1}); brain: -48.5 kJ mol^{-1} (-11.6 kcal mol^{-1}). The ΔG is most negative in brain cells.

24. (a) Ethanol; (b) lactate; (c) succinate; (d) isocitrate; (e) malate.

25. Recall that $\Delta G = \Delta G^{\circ\prime} + RT \ln [\text{products}]/[\text{reactants}]$. Altering the ratio of products to reactants will cause ΔG to vary. In glycolysis, the concentrations of the components of the pathway result in a value of ΔG greater than that of $\Delta G^{\circ\prime}$.

26. Higher organisms cannot make vitamins, and thus are dependent on obtaining them from other organisms.

27. Unless the ingested food is converted into molecules capable of being absorbed by the intestine, no energy can ever be extracted by the body.

28. NADH and $FADH_2$ are electron carriers for catabolism; NADPH is the carrier for anabolism.

29. The electrons of the C–O bond cannot form resonance structures with the C–S bond that are as stable as those that they can form with the C–O bond. Thus, the thioester is not stabilized by resonance to the same degree as an oxygen ester is stabilized.

30. Oxidation–reduction reactions; ligation reactions; isomerization reactions; group-transfer reactions; hydrolytic reactions; cleavage of bonds by means other than hydrolysis or oxidation.

31. Controlling the amount of enzymes; controlling enzyme activity; controlling the availability of substrates.

32. Although the reaction is thermodynamically favorable, the reactants are kinetically stable because of the large activation energy. Enzymes lower the activation energy so that reactions take place on time scales required by the cell.

33. The activated form of sulfate in most organisms is 3′-phosphoadenosine-5′-phosphosulfate.

34. (a) As the Mg^{2+} concentration falls, the ΔG of hydrolysis rises. Note that pMg is a logarithmic plot, and so each number on the x-axis represents a 10-fold change in $[Mg^{2+}]$.
(b) Mg^{2+} would bind to the phosphates of ATP and help to mitigate charge repulsion. As the $[Mg^{2+}]$ falls, charge stabilization of ATP would be less, leading to greater charge repulsion and an increase in ΔG on hydrolysis.

Chapter 16

1. Two molecules of ATP are produced per molecule of glyceraldehyde 3-phosphate and, because two molecules of GAP are produced per molecule of glucose, the total ATP yield is four. However, two molecules of ATP are required to convert glucose into fructose 1,6-bisphosphate. Thus, the net yield is only two molecules of ATP.

2. (a) 4; (b) 3; (c) 1; (d) 6; (e) 8; (f) 2; (g) 10; (h) 9; (i) 7; (j) 5.

3. In both cases, the electron donor is glyceraldehyde 3-phosphate. In lactic acid fermentation, the electron acceptor is pyruvate, converting it into lactate. In alcoholic fermentation, acetaldehyde is the electron acceptor, forming ethanol.

4. (a) 3 ATP; (b) 2 ATP; (c) 2 ATP; (d) 2 ATP; (e) 4 ATP.

5. Glucokinase enables the liver to remove glucose from the blood when hexokinase is saturated, ensuring that glucose is captured for later use.

6. The GAP formed is immediately removed by subsequent reactions, resulting in the conversion of DHAP into GAP by the enzyme.

7. A thioester couples the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate with the formation of 1,3-bisphosphoglycerate. 1,3-Bisphosphoglycerate can subsequently power the formation of ATP.

8. Glycolysis is a component of alcoholic fermentation, the pathway that produces alcohol for beer and wine. The belief was that understanding the biochemical basis of alcohol production might lead to a more-efficient means of producing beer.

9. The conversion of glyceraldehyde 3-phosphate into 1,3- bisphosphoglycerate would be impaired. Glycolysis would be less effective.

10. Glucose 6-phosphate must have other fates. Indeed, it can be converted into glycogen (Chapter 21) or be processed to yield reducing power for biosynthesis (Chapter 20).

11. The energy needs of a muscle cell vary widely, from rest to intense exercise. Consequently, the regulation of phosphofructokinase by energy charge is vital. In other tissues, such as the liver, ATP concentration is less likely to fluctuate and will not be a key regulator of phosphofructokinase.

12. (a) 6; (b) 1; (c) 7; (d) 3; (e) 2; (f) 5 ; (g) 4.

13. The $\Delta G^{\circ\prime}$ for the reverse of glycolysis is $+90$ kJ mol^{-1} ($+22$ kcal mol^{-1}), far too endergonic to take place.

14. The conversion of glucose into glucose 6-phosphate by hexokinase; the conversion of fructose 6-phosphate into fructose 1,6-bisphosphate by phosphofructokinase; the formation of pyruvate from phosphoenolpyruvate by pyruvate kinase.

15. Lactic acid is a strong acid. If it remained in the cell, the pH of the cell would fall, which could lead to the denaturation of muscle protein and result in muscle damage.

16. GLUT2 transports glucose only when the blood concentration of glucose is high, which is precisely the condition in which the β cells of the pancreas secrete insulin.

17. Fructose + ATP $\longrightarrow$ fructose 1-phosphate + ADP: Fructokinase

Fructose 1-phosphate $\longrightarrow$ dihydroxyacetone phosphate + glyceraldehyde: Fructose 1-phosphate aldolase

Glyceraldehyde + ATP $\longrightarrow$ glyceraldehyde 3-phosphate + ADP: Triose kinase

The primary controlling step of glycolysis catalyzed by phosphofructokinase is bypassed by the preceding reactions. Glycolysis will proceed in an unregulated fashion.

18. (a) A, B; (b) C, D; (c) D; (d) A; (e) B; (f) C; (g) A; (h) D; (i) none; (j) A; (k) A.

19. (a) 4; (b) 10; (c) 1; (d) 5; (e) 7; (f) 8; (g) 9; (h) 2; (i) 3; (j) 6.

20. Without triose isomerase, only one of the two three-carbon molecules generated by aldolase could be used to generate ATP. Only two molecules of ATP would result from the metabolism of each glucose. But two molecules of ATP would still be required to form fructose 1,6-bisphosphate, the substrate for aldolase. The net yield of ATP would be zero, a yield incompatible with life.

21. Glucose is reactive because its open-chain form contains an aldehyde group.

22. (a) The label is in the methyl carbon atom of pyruvate. (b) 5 mCi mM^{-1}. The specific activity is halved because the number of moles of product (pyruvate) is twice that of the labeled substrate (glucose).

23. (a) Glucose + 2 P_i + 2 ADP $\rightarrow$ 2 lactate + 2 ATP.
(b) $\Delta G = -114$ kJ mol^{-1} (-27.2 kcal mol^{-1}).

24. 3.06×10^{-5}

25. The equilibrium concentrations of fructose 1,6-bisphosphate, dihydroxyacetone phosphate, and glyceraldehyde 3-phosphate are 7.8×10^{-4} M, 2.2×10^{-4} M, and 2.2×10^{-4} M, respectively.

26. All three carbon atoms of 2,3-BPG are ^{14}C labeled. The phosphorus atom attached to the C-2 hydroxyl group is ^{32}P labeled.

27. Hexokinase has a low ATPase activity in the absence of a sugar because it is in a catalytically inactive conformation. The addition of xylose closes the cleft between the two lobes of the enzyme. However, xylose lacks a hydroxymethyl group, and so it cannot be phosphorylated. Instead, a water molecule at the site normally occupied by the C-6 hydroxymethyl group acts as the acceptor of the phosphoryl group from ATP.
28. (a) The fructose 1-phosphate pathway forms glyceraldehyde 3-phosphate.
(b) Phosphofructokinase, a key control enzyme, is bypassed.
29. The reverse of glycolysis is highly endergonic under cellular conditions. The expenditure of six NTP molecules in gluconeogenesis renders gluconeogenesis exergonic.
30. (a) 2, 3, 6, 9; (b) 1, 4, 5, 7, 8.
31. Lactic acid is capable of being further oxidized and is thus useful energy. The conversion of this acid into glucose saves the carbon atoms for future combustion.
32. In glycolysis, the formation of pyruvate and ATP by pyruvate kinase is irreversible. This step is bypassed by two reactions in gluconeogenesis: (1) the formation of oxaloacetate from pyruvate and CO_2 by pyruvate carboxylase and (2) the formation of phosphoenolpyruvate from oxaloacetate and GTP by phosphoenolpyruvate carboxykinase. The formation of fructose 1,6-bisphosphate by phosphofructokinase is bypassed by fructose 1,6-bisphosphatase in gluconeogenesis, which catalyzes the conversion of fructose 1,6-bisphosphate into fructose 6-phosphate. Finally, the hexokinase-catalyzed formation of glucose 6-phosphate in glycolysis is bypassed by glucose 6-phosphatase, but only in the liver.
33. Reciprocal regulation at the key allosteric enzymes in the two pathways. For instance, PFK is stimulated by fructose 2,6-bisphosphate and AMP. The effect of these signals is opposite that of fructose 1,6-bisphosphatase. If both pathways were operating simultaneously, a futile cycle would result. ATP would be hydrolyzed, yielding only heat.
34. Muscle is likely to produce lactic acid during contraction. Lactic acid is a strong acid and cannot accumulate in muscle or blood. Liver removes the lactic acid from the blood and converts it into glucose. The glucose can be released into the blood or stored as glycogen for later use.
35. Glucose produced by the liver could not be released into the blood. Tissues that rely on glucose as an energy source would not function as well unless glucose was provided in the diet.
36. Glucose is an important energy source for both tissues and is essentially the only energy source for the brain. Consequently, these tissues should never release glucose. Glucose release is prevented by the absence of glucose 6-phosphatase.
37. 6 NTP (4 ATP and 2 GTP); 2 NADH.
38. (a) None; (b) none; (c) 4 (2 ATP and 2 GTP); (d) none.
39. If the amino groups are removed from alanine and aspartate, the ketoacids pyruvate and oxaloacetate are formed. Both of these molecules are components of the gluconeogenic pathway.
40. (a) Increased; (b) increased; (c) increased; (d) decreased.
41. Fructose 2,6-bisphosphate, present at high concentration when glucose is abundant, normally inhibits gluconeogenesis by blocking fructose 1,6-bisphosphatase. In this genetic disorder, the phosphatase is active irrespective of the glucose level. Hence, substrate cycling is increased. The level of fructose 1,6-bisphosphate is consequently lower than normal. Less pyruvate is formed and thus less ATP is generated.
42. Reactions in parts *b* and *e* would be blocked.
43. There will be no labeled carbons. The CO_2 added to pyruvate (formed from the lactate) to form oxaloacetate is lost with the conversion of oxaloacetate into phosphoenolpyruvate.
44. The net reaction in the presence of arsenate is

$$\text{Glyceraldehyde 3-phosphate} + \text{NAD}^+ + \text{H}_2\text{O} \longrightarrow \text{3-phosphoglycerate} + \text{NADH} + 2\,\text{H}^+$$

Glycolysis proceeds in the presence of arsenate, but the ATP normally formed in the conversion of 1,3-bisphosphoglycerate into 3-phosphoglycerate is lost. Thus, arsenate uncouples oxidation and phosphorylation by forming a highly labile acyl arsenate.
45. This example illustrates the difference between the *stoichiometric* and the *catalytic* use of a molecule. If cells used NAD^+ stoichiometrically, a new molecule of NAD^+ would be required each time a molecule of lactate was produced. As we will see, the synthesis of NAD^+ requires ATP. On the other hand, if the NAD^+ that is converted into NADH could be recycled and reused, a small amount of the molecule could regenerate a vast amount of lactate, which is the case in the cell. NAD^+ is regenerated by the oxidation of NADH and reused. NAD^+ is thus used catalytically.
46. Consider the equilibrium equation of adenylate kinase:

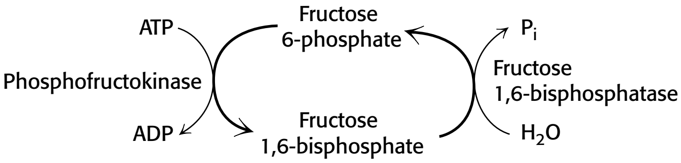
$$K_{\text{eq}} = [\text{ATP}][\text{AMP}]/\,[\text{ADP}]^2 \,(1)$$

or

$$\text{AMP} = K_{\text{eq}}\,[\text{ADP}]^2/[\text{ATP}](2)$$

Recall that [ATP] > [ADP] > [AMP] in the cell. As ATP is utilized, a small decrease in its concentration will result in a larger percentage increase in [ADP] because its concentration is greater than that of ADP. This larger percentage increase in [ADP] will result in an even greater percentage increase in [AMP] because the concentration of AMP is related to the square of [ADP]. In essence, equation 2 shows that monitoring the energy status with AMP magnifies small changes in [ATP], leading to tighter control.
47. The synthesis of glucose during intense exercise provides a good example of interorgan cooperation in higher organisms. When muscle is actively contracting, lactate is produced from glucose by glycolysis. The lactate is released into the blood and absorbed by the liver, where it is converted by gluconeogenesis into glucose. The newly synthesized glucose is then released and taken up by the muscle for energy generation.
48. The input of four additional high-phosphoryl-transfer-potential molecules in gluconeogenesis changes the equilibrium constant by a factor of 10^{32}, which makes the conversion of pyruvate into glucose thermodynamically feasible. Without this energetic input, gluconeogenesis would not take place.
49. The mechanism is analogous to that for triose phosphate isomerase (Figure 16.5). It proceeds through an enediol intermediate. The active site would be expected to have a general base (analogous to Glu 165 in TPI) and a general acid (analogous to His 95 in TPI).
50. Galactose is a component of glycoproteins. Possibly, the absence of galactose leads to the improper formation or function of glycoproteins required in the central nervous system. More generally, the fact that the symptoms arise in the absence of galactose suggests that galactose is required in some fashion.
51. Using the Michaelis–Menten equation to solve for [S] when $K_M = 50\ \mu M$ and $V_o = 0.9V_{max}$ shows that a substrate concentration of 0.45 mM yields 90% of V_{max}. Under normal conditions, the enzyme is essentially working at V_{max}.
52. Fructose 2,6-bisphosphate stabilizes the R state of the enzyme.
53. (a) Curiously, the enzyme uses ADP as the phosphoryl donor rather than ATP.
(b) Both AMP and ATP behave as competitive inhibitors of ADP, the phosphoryl donor. Apparently, the *P. furiosus* enzyme is not allosterically inhibited by ATP.

54. (a) If both enzymes operated simultaneously, the following reactions would take place:



The net result would be simply:

$$\text{ATP} + H_2O \longrightarrow \text{ADP} + P_i$$

The energy of ATP hydrolysis would be released as heat.
(b) Not really. For the cycle to generate heat, both enzymes must be functional at the same time in the same cell.
(c) The species *B. terrestris* and *B. rufocinctus* might show some futile cycling because both enzymes are active to a substantial degree.
(d) No. These results simply suggest that simultaneous activity of phosphofructokinase and fructose 1,6-bisphosphatase is unlikely to be employed to generate heat in the species shown.
55. ATP initially stimulates PFK activity, as would be expected for a substrate. Higher concentrations of ATP inhibit the enzyme. Although this effect seems counterintuitive for a substrate, recall that the function of glycolysis in muscle is to generate ATP. Consequently, high concentrations of ATP signal that the ATP needs are met and glycolysis should stop. In addition to being a substrate, ATP is an allosteric inhibitor of PFK.

Chapter 17

1. The pyruvate dehydrogenase complex catalyzes the following reaction, linking glycolysis and the citric acid cycle:

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \longrightarrow \text{acetyl CoA} + \text{NADH} + H^+ + CO_2$$

2. Pyruvate dehydrogenase catalyzes the decarboxylation of pyruvate and the formation of acetyllipoamide. Dihydrolipoyl transacetylase catalyzes the formation of acetyl CoA. Dihydrolipoyl dehydrogenase catalyzes the reduction of the oxidized lipoic acid. The kinase associated with the complex phosphorylates and inactivates the complex, whereas the phosphatase dephosphorylates and activates the complex.
3. Thiamine pyrophosphate plays a role in the decarboxylation of pyruvate. Lipoic acid (as lipoamide) transfers the acetyl group. Coenzyme A accepts the acetyl group from lipoic acid to form acetyl CoA. FAD accepts the electrons and hydrogen ions when reduced lipoic acid is oxidized. NAD^+ accepts electrons from FADH_2.
4. Catalytic coenzymes (TPP, lipoic acid, and FAD) are modified but regenerated in each reaction cycle. Thus, they can play a role in the processing of many molecules of pyruvate. Stoichiometric coenzymes (coenzyme A and NAD^+) are used in only one reaction because they are the components of products of the reaction.
5. The remaining steps regenerate oxidized lipoamide, which is required to begin the next reaction cycle. Moreover, this regeneration results in the production of high-energy electrons in the form of NADH.
6. The advantages are as follows:
   The reaction is facilitated by having the active sites in proximity.
   The reactants do not leave the enzyme until the final product is formed.
   Constraining the reactants minimizes loss due to diffusion and minimizes side reactions.
   All of the enzymes are present in the correct amounts.
   Regulation is more efficient because the regulatory enzymes—the kinase and phosphatase—are part of the complex.
7. (a) After one round of the citric acid cycle, the label emerges in C-2 and C-3 of oxaloacetate. (b) The label emerges in CO_2 in the formation of acetyl CoA from pyruvate. (c) After one round of the citric acid cycle, the label emerges in C-1 and C-4 of oxaloacetate. (d) and (e) Same fate as that in part *a*.
8. (a) Isocitrate lyase and malate synthase are required in addition to the enzymes of the citric acid cycle.
(b) 2 Acetyl CoA + 2 NAD^+ + FAD + 3 $H_2O \longrightarrow$ oxaloacetate + 2 CoA + 2 NADH + FADH_2 + 3 H^+.
(c) No. Hence, mammals cannot carry out the net synthesis of oxaloacetate from acetyl CoA.
9. $-41.0 \text{ kJ mol}^{-1} (-9.8 \text{ kcal mol}^{-1})$
10. Enzymes or enzyme complexes are biological catalysts. Recall that a catalyst facilitates a chemical reaction without the catalyst itself being permanently altered. Oxaloacetate can be thought of as a catalyst because it binds to an acetyl group, leads to the oxidative decarboxylation of the two carbon atoms, and is regenerated at the completion of a cycle. In essence, oxaloacetate (and any cycle intermediate) acts as a catalyst.
11. Thiamine thiazolone pyrophosphate is a transition-state analog. The sulfur-containing ring of this analog is uncharged, and so it closely resembles the transition state of the normal coenzyme in thiamine-catalyzed reactions (e.g., the uncharged resonance form of hydroxyethyl-TPP).
12. (a) 6; (b) 10; (c) 1; (d) 7; (e) 2; (f) 8; (g) 3; (h) 4; (i) 5; (j) 9.
13. A decrease in the amount of O_2 will necessitate an increase in anaerobic glycolysis for energy production, leading to the generation of a large amount of lactic acid. Under conditions of shock, the kinase inhibitor is administered to ensure that pyruvate dehydrogenase is operating maximally.
14. (a) As is stated in the previous problem, DCA inhibits pyruvate dehydrogenase kinase.
(b) The fact that inhibiting the kinase results in more dehydrogenase activity suggests that there must be some residual activity that is being inhibited by the kinase.
15. Acetyllipoamide and acetyl CoA.
16. In muscle, the acetyl CoA generated by the complex is used for energy generation. Consequently, signals that indicate an energy-rich state (high ratios of ATP/ADP and NADH/NAD^+) inhibit the complex, whereas the reverse conditions stimulate the enzyme. Calcium as the signal for muscle contraction (and, hence, energy need) also stimulates the enzyme. In liver, acetyl CoA derived from pyruvate is used for biosynthetic purposes, such as fatty acid synthesis. Insulin, the hormone denoting the fed state, stimulates the complex.
17. (a) Enhanced kinase activity will result in a decrease in the activity of the PDH complex because phosphorylation by the kinase inhibits the complex.
(b) Phosphatase activates the complex by removing a phosphate. If the phosphatase activity is diminished, the activity of the PDH complex also will decrease.
18. She might have been ingesting, in some fashion, the arsenite from the peeling paint or the wallpaper. Also, she might have been breathing arsine gas from the wallpaper, which would be oxidized to arsenite in her body. In any of these circumstances, the arsenite inhibited enzymes that require lipoic acid—notably, the PDH complex.
19. (a) 5; (b) 7; (c) 1; (d) 10; (e) 2; (f) 4; (g) 9; (h) 3; (i) 8; (j) 6.
20. The TCA cycle depends on a steady supply of NAD^+ as an oxidant, generating NADH. O_2 is never directly utilized in the cycle. However, NAD^+ is regenerated via donation of electrons to O_2 by way of the electron transport chain, so eventually a lack of O_2 will cause the cycle to cease due to a lack of NAD^+.

21. Succinate dehydrogenase is the only enzyme in the citric acid cycle that is embedded in the mitochondrial membrane, which makes it associated with the electron-transport chain.

22. (a) The steady-state concentrations of the products are low compared with those of the substrates. (b) The ratio of malate to oxaloacetate must be greater than 1.57×10^4 for oxaloacetate to be formed.

23.

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \xrightarrow{\substack{\text{Pyruvate}\\ \text{dehydrogenase}\\ \text{complex}}} \text{acetyl CoA} + \text{CO}_2 + \text{NADH}$$

$$\text{Pyruvate} + \text{CO}_2 + \text{ATP} + \text{H}_2\text{O} \xrightarrow{\substack{\text{Pyruvate}\\ \text{carboxylase}}} \text{oxaloacetate} + \text{ADP} + \text{P}_i + \text{H}^+$$

$$\text{Oxaloacetate} + \text{acetyl CoA} + \text{H}_2\text{O} \xrightarrow{\substack{\text{Citrate}\\ \text{synthase}}} \text{citrate} + \text{CoA} + \text{H}^+$$

$$\text{Citrate} \xrightarrow{\text{Aconitase}} \text{isocitrate}$$

$$\text{Isocitrate} + \text{NAD}^+ \xrightarrow{\substack{\text{Isocitrate}\\ \text{dehydrogenase}}} \alpha\text{-ketoglutarate} + \text{CO}_2 + \text{NADH}$$

$$\text{Net: } 2\ \text{Pyruvate} + 2\ \text{NAD}^+ + \text{ATP} + \text{H}_2\text{O} \longrightarrow \alpha\text{-ketoglutarate} + \text{CO}_2 + \text{ADP} + \text{P}_i + 2\ \text{NADH} + 3\ \text{H}^+$$

24. Succinate will increase in concentration, followed by α-ketoglutarate and the other intermediates "upstream" of the site of inhibition. Succinate has two methylene groups that are required for the dehydrogenation, whereas malonate has but one.

25. Pyruvate carboxylase should be active only when the acetyl CoA concentration is high. Acetyl CoA might accumulate if the energy needs of the cell are not being met, because of a deficiency of oxaloacetate. Under these conditions the pyruvate carboxylase catalyzes an anaplerotic reaction. Alternatively, acetyl CoA might accumulate because the energy needs of the cell have been met. In this circumstance, pyruvate will be converted back into glucose, and the first step in this conversion is the formation of oxaloacetate.

26. The energy released when succinate is oxidized to fumarate is not sufficient to power the synthesis of NADH but is sufficient to reduce FAD.

27. Citrate is a tertiary alcohol that cannot be oxidized, because oxidation requires a hydrogen atom to be removed from the alcohol and a hydrogen atom to be removed from the carbon atom bonded to the alcohol. No such hydrogen exists in citrate. The isomerization converts the tertiary alcohol into isocitrate, which is a secondary alcohol that can be oxidized.

28. The enzyme nucleoside diphosphokinase transfers a phosphoryl group from GTP (or any nucleoside triphosphate) to ADP according to the reversible reaction:

$$\text{GTP} + \text{ADP} \rightleftharpoons \text{GDP} + \text{ATP}$$

29. The reaction is powered by the hydrolysis of a thioester. Acetyl CoA provides the thioester that is converted into citryl CoA. When this thioester is hydrolyzed, citrate is formed in an irreversible reaction.

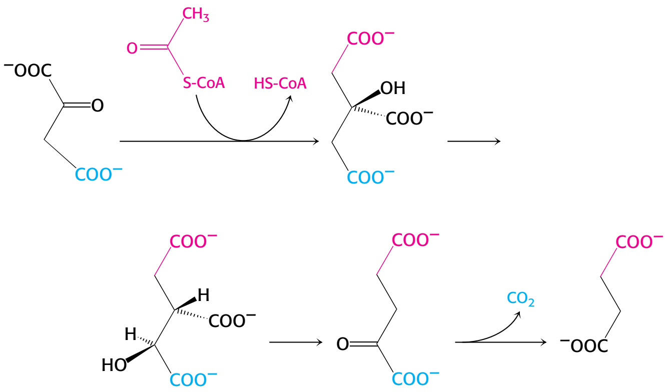
30. It enables organisms such as plants and bacteria to convert fats, through acetyl CoA, into glucose.

31. We cannot get the net conversion of fats into glucose, because the only means to get the carbon atoms from fats into oxaloacetate, the precursor of glucose, is through the citric acid cycle. However, although two carbon atoms enter the cycle as acetyl CoA, two carbon atoms are lost as CO_2 before oxaloacetate is formed. Thus, although some carbon atoms from fats may end up as carbon atoms in glucose, we cannot obtain a *net* synthesis of glucose from fats.

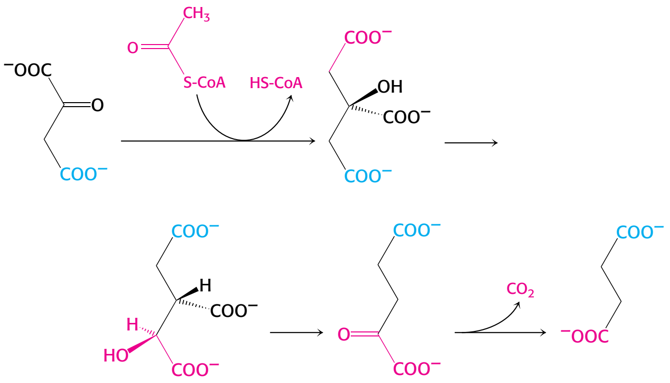
32. Acetyl CoA will inhibit the complex. Glucose metabolism to pyruvate will be slowed because acetyl CoA is being derived from an alternative source.

33. The enol intermediate of acetyl CoA attacks the carbonyl carbon atom of glyoxylate to form a C–C bond. This reaction is like the condensation of oxaloacetate with the enol intermediate of acetyl CoA in the reaction catalyzed by citrate synthase. Glyoxylate contains a hydrogen atom in place of the $-\text{CH}_2\text{COO}^-$ group of oxaloacetate; the reactions are otherwise nearly identical.

34. Citrate is a symmetric molecule. Consequently, the investigators assumed that the two $-\text{CH}_2\text{COO}^-$ groups in it would react identically. Thus, for every citrate molecule undergoing the reactions shown in path 1, they thought that another citrate molecule would react as shown in path 2. If so, then only *half* the label should have emerged in the CO_2.

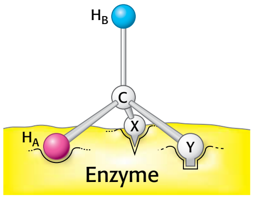


Path 1



Path 2 (does not occur)

35. Call one hydrogen atom A and the other B. Now suppose that an enzyme binds three groups of this substrate—X, Y, and H—at three complementary sites. The adjoining diagram shows X, Y, and H_A bound to three points on the enzyme. In contrast, X, Y, and H_B cannot be bound to this active site; two of these three groups can be bound, but not all three. Thus, H_A and H_B will have different fates.



Sterically nonequivalent groups such as H_A and H_B will almost always be distinguished in enzymatic reactions. The essence of the differentiation of these groups is that the enzyme holds the substrate in a specific orientation. Attachment at three points, as depicted in the diagram, is a readily visualized way of achieving a particular orientation of the substrate, but it is not the only means of doing so.

36. (a) The complete oxidation of citrate requires 4.5 μmol of O_2 for every micromole of citrate.

$$C_6H_8O_7 + 4.5\ O_2 \longrightarrow 6\ CO_2 + 4\ H_2O$$

Thus, 13.5 μmol of O_2 would be consumed by 3 μmol of citrate. (b) Citrate led to the consumption of far more O_2 than can be accounted for simply by the oxidation of citrate itself. Citrate thus facilitated O_2 consumption.

37. (a) In the absence of arsenite, the amount of citrate remained constant. In its presence, the concentration of citrate fell, suggesting that it was being metabolized.
(b) The action of arsenite is not altered. Citrate still disappears.
(c) Arsenite is preventing the regeneration of citrate. Recall (pp. 515–516) that arsenite inhibits the pyruvate dehydrogenase complex.

38. (a) The initial infection is unaffected by the absence of isocitrate lyase, but the absence of this enzyme inhibits the latent phase of the infection.
(b) Yes.
(c) A critic could say that, in the process of deleting the isocitrate lyase gene, some other gene was damaged, and it is the absence of this other gene that prevents latent infection. Reinserting the isocitrate lyase gene into the bacteria from which it had been removed renders the criticism less valid.
(d) Isocitrate lyase enables the bacteria to synthesize carbohydrates that are necessary for survival, including carbohydrate components of the cell membrane.

Chapter 18

1. In fermentations, organic compounds are both the donors and the acceptors of electrons. In respiration, the electron donor is usually an organic compound, whereas the electron acceptor is an inorganic molecule, such as oxygen.

2. Biochemists use E'_0, the value at pH 7, whereas chemists use E_0, the value in 1 M H^+. The prime denotes that pH 7 is the standard state.

3. The reduction potential of $FADH_2$ is less than that of NADH (Table 18.1). Consequently, when those electrons are passed along to oxygen, less energy is released. The consequence of the difference is that electron flow from $FADH_2$ to O_2 pumps fewer protons than do the electrons from NADH.

4. The $\Delta G^{\circ\prime}$ for the reduction of oxygen by $FADH_2$ is -200 kJ mol^{-1} (-48 kcal mol^{-1}).

5. $\Delta G^{\circ\prime}$ is $+67$ kJ mol^{-1} ($+16.1$ kcal mol^{-1}) for oxidation by NAD^+ and -3.8 kJ mol^{-1} (-0.92 kcal mol^{-1}) for oxidation by FAD. The oxidation of succinate by NAD^+ is not thermodynamically feasible.

6. An oxidizing agent, or oxidant, accepts electrons in oxidation–reduction reactions. A reducing reagent, or reductant, donates electrons in such reactions.

7. Pyruvate accepts electrons and is thus the oxidant. NADH gives up electrons and is the reductant.

8. $\Delta G^{\circ\prime} = -nF\Delta E'_0$

9. The $\Delta E'_0$ value of iron can be altered by changing the environment of the ion.

10. c, e, b, a, d.

11. (a) 4; (b) 5; (c) 2; (d) 10; (e) 3; (f) 8; (g) 9; (h) 7; (i) 1; (j) 6.

12. (a) 4; (b) 3; (c) 1; (d) 5; (e) 2.

13. The 10 isoprene units render coenzyme Q soluble in the hydrophobic environment of the inner mitochondrial membrane. The two oxygen atoms can reversibly bind two electrons and two protons as the molecule transitions between the quinone form and quinol form.

14. Rotenone: NADH, NADH-Q oxidoreductase will be reduced. The remainder will be oxidized. Antimycin A: NADH, NADH-Q oxidoreductase and coenzyme Q will be reduced. The remainder will be oxidized. Cyanide: All will be reduced.

15. Complex I would be reduced, whereas Complexes II, III, and IV would be oxidized. The citric acid cycle would halt because it has no way to oxidize NADH.

16. The respirasome is another example of the use of supramolecular complexes in biochemistry. Having the three complexes that are proton pumps associated with one another will enhance the efficiency of electron flow from complex to complex, which in turn will cause more-efficient proton pumping.

17. Succinate dehydrogenase is a component of Complex II.

18. Hydroxyl radical ($OH\cdot$), hydrogen peroxide (H_2O_2), superoxide ion ($O_2^{\cdot -}$), and peroxide (O_2^{2-}). These small molecules react with a host of macromolecules—including proteins, nucleotides, and membranes—to disrupt cell structure and function.

19. The ATP is recycled by ATP-generating processes, most notably oxidative phosphorylation.

20. (a) 12.5; (b) 14; (c) 32; (d) 13.5; (e) 30; (f) 16.

21. (a) It blocks electron transport and proton pumping at Complex IV. (b) It blocks electron transport and ATP synthesis by inhibiting the exchange of ATP and ADP across the inner mitochondrial membrane. (c) It blocks electron transport and proton pumping at Complex I. (d) It blocks ATP synthesis without inhibiting electron transport by dissipating the proton gradient. (e) It blocks electron transport and proton pumping at Complex IV. (f) It blocks electron transport and proton pumping at Complex III.

22. If the proton gradient is not dissipated by the influx of protons into a mitochondrion with the generation of ATP, eventually the outside of the mitochondrion develops such a large positive charge that the electron-transport chain can no longer pump protons against the gradient.

23. The subunits are jostled by background thermal energy (Brownian motion). The proton gradient makes clockwise rotation more likely because that direction results in protons flowing down their concentration gradient.

24. Dicyclohexylcarbodiimide reacts readily with carboxyl groups. Hence, the most likely targets are aspartate and glutamate side chains. In fact, Asp 61 of subunit **c** of *E. coli* F_0 is specifically modified by this reagent. The conversion of Asp 61 into asparagine by site-specific mutagenesis eliminates proton conduction.

25. In the presence of poorly functioning mitochondria, the only means of generating ATP is by anaerobic glycolysis, which will lead to an accumulation of lactic acid in blood.

26. If ADP cannot get into mitochondria, the electron-transport chain will cease to function because there will be no acceptor for

the energy. NADH will build up in the matrix. Recall that NADH inhibits some citric acid cycle enzymes and that NAD^+ is required by several citric acid cycle enzymes. Glycolysis will switch to lactic acid fermenation so that the NADH can be reoxidized to NAD^+ by lactate dehydrogenase.

27. (a) No effect; mitochondria cannot metabolize glucose.
(b) No effect; no fuel is present to power the synthesis of ATP.
(c) The $[O_2]$ falls because citrate is a fuel and ATP can be formed from ADP and P_i.
(d) Oxygen consumption stops because oligomycin inhibits ATP synthesis, which is coupled to the activity of the electron-transport chain.
(e) No effect, for the reasons given in part *d*.
(f) $[O_2]$ falls rapidly because the system is uncoupled and does not require ATP synthesis to lower the proton-motive force.
(g) $[O_2]$ falls, though at a lower rate. Rotenone inhibits Complex I, but the presence of succinate will enable electrons to enter at Complex II.
(h) Oxygen consumption ceases because Complex IV is inhibited and the entire chain backs up.

28. (a) The P : O ratio is equal to the product of $(H^+/2\ e^-)$ and (P/H^+). Note that the P : O ratio is identical with the P : 2 e^- ratio.
(b) 2.5 and 1.5, respectively.

29. Cyanide can be lethal because it binds to the ferric form of cytochrome oxidase and thereby inhibits oxidative phosphorylation. Nitrite converts ferrohemoglobin into ferrihemoglobin, which also binds cyanide. Thus, ferrihemoglobin competes with cytochrome *c* oxidase for cyanide. This competition is therapeutically effective because the amount of ferrihemoglobin that can be formed without impairing oxygen transport is much greater than the amount of cytochrome *c* oxidase.

30. Such a defect (called Luft syndrome) was found in a 38-year-old woman who was incapable of performing prolonged physical work. Her basal metabolic rate was more than twice normal, but her thyroid function was normal. A muscle biopsy showed that her mitochondria were highly variable and atypical in structure. Biochemical studies then revealed that oxidation and phosphorylation were not tightly coupled in these mitochondria. In this patient, much of the energy of fuel molecules was converted into heat rather than ATP.

31. Triose phosphate isomerase converts dihydroxyacetone phosphate (a potential dead end) into glyceraldehyde 3-phosphate (a mainstream glycolytic intermediate).

32. This inhibitor (like antimycin A) blocks the reduction of cytochrome c_1 by QH_2, the crossover point.

33. If oxidative phosphorylation were uncoupled, no ATP could be produced. In a futile attempt to generate ATP, much fuel would be consumed. The danger lies in the dose. Too much uncoupling would lead to tissue damage in highly aerobic organs such as the brain and heart, which would have severe consequences for the organism as a whole. The energy that is normally transformed into ATP would be released as heat. To maintain body temperature, sweating might increase, although the very process of sweating itself depends on ATP.

34. If ATP and ADP cannot exchange between the matrix and the mitochondria, ATP synthase will cease to function because its substrate ADP is absent. The proton gradient will eventually become so large that the energy released by the electron-transport chain will not be great enough to pump protons against the larger-than-normal gradient.

35. Add the inhibitor with and without an uncoupler, and monitor the rate of O_2 consumption. If the O_2 consumption increases again in the presence of inhibitor and uncoupler, the inhibitor must be inhibiting ATP synthase. If the uncoupler has no effect on the inhibition, the inhibitor is inhibiting the electron-transport chain.

36. Presumably, because the muscle has greater energy needs, especially during exercise, it will require more ATP. This requirement means that more sites of oxidative phosphorylation are called for, and these sites can be provided by an increase in the amount of cristae.

37. The arginine residue, with its positive charge, will facilitate proton release from aspartic acid by stabilizing the negatively charged aspartate.

38. 4; 4.7

39. The ATP synthase would pump protons at the expense of ATP hydrolysis, thus maintaining the proton-motive force. The synthase would function as an ATPase. There is some evidence that damaged mitochondria use this tactic to maintain, at least temporarily, the proton-motive force.

40. It suggests that malfunctioning mitochondria may play a role in the development of Parkinson disease. Specifically, it implicates Complex I.

41. The extra negative charge on ATP relative to that on ADP accounts for ATP's more-rapid translocation out of the mitochondrial matrix. If the charge differences between ATP and ADP were lessened by the binding of Mg^{2+}, ADP might more readily compete with ATP for transport to the cytoplasm.

42. When all of the available ADP has been converted into ATP, ATP synthase can no longer function. The proton gradient becomes large enough that the energy of the electron-transport chain is not enough to pump against the gradient, and electron transport and, hence, oxygen consumption falls.

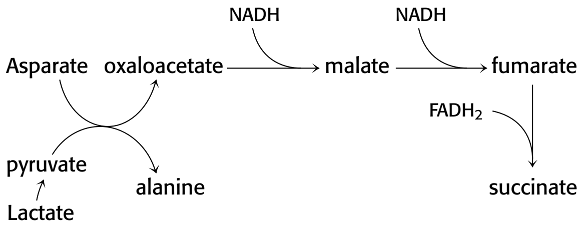
43. The effect on the proton gradient is the same in each case.

44 ATP export from the matrix. Phosphate import into the matrix.

45. Recall from the discussion of enzyme-catalyzed reactions that the direction of a reaction is determined by the ΔG difference between substrate and products. An enzyme speeds up the rate of both the forward and the backward reactions. The hydrolysis of ATP is exergonic, and so ATP synthase will enhance the hydrolytic reaction.

46. The cytoplasmic kinases thereby obtain preferential access to the exported ATP.

47. The organic acids in the blood are indications that the mice are deriving a large part of their energy needs through anaerobic glycolysis. Lactate is the end product of anaerobic glycolysis. Alanine is an aminated transport form of pyruvate, which is formed from lactate. Alanine formation plays a role in succinate formation, which is caused by the reduced state of the mitochondria.



The electron-transport chain is slowed because the inner mitochondrial membrane is hyperpolarized. Without ADP to accept the energy of the proton-motive force, the membrane becomes polarized to such an extent that protons can no longer be pumped. The excess H_2O_2 is probably due to the fact that the superoxide radical is present in higher concentration because the oxygen can no longer be effectively reduced.

$$O_2\cdot^- + O_2\cdot^- + 2\,H^+ \longrightarrow O_2 + H_2O_2$$

Indeed, these mice display evidence of such oxidative damage.

48. (a) Vitamins C and E.
(b) Exercise induces superoxide dismutase, which converts ROS in hydrogen peroxide and oxygen.

(c) The answer to this question is not fully established. Two possibilities are (1) the suppression of ROS by vitamins prevents the expression of more superoxide dismutase and (2) some ROS may be signal molecules required to stimulate insulin-sensitivity pathways.

49. (a) DNP is an uncoupler that prevents the use of the proton-motive force for ATP synthesis. Consequently, the rate of oxygen consumption rises (reflecting the rate of the electron transport chain) in a futile attempt to synthesize ATP. Because mitochondrial ATP synthesis in inhibited, the rate of glycolysis, as measured by ECAR, increases in an attempt meet the cells ATP needs.
(b) Because glycolysis is now inhibited, no lactic acid will be produced and the rate of extracellular acidification will fall. Because DNP is still present, oxygen consumption will still occur at a high rate.
(c) A key step in glycolysis is the isomerization of glucose 6-phosphate to fructose 6-phosphate. 2-Deoxyglucose is incapable of undergoing the isomerization.
(d) Rotenone inhibits electron flow through Complex I, the electron transport chain is inhibited and oxygen consumption ceases.

50. (a) Succinate is oxidized by Complex II, and the electrons are used to establish a proton-motive force that powers ATP synthesis.
(b) The ability to synthesize ATP is greatly reduced.
(c) The goal was to measure ATP hydrolysis. If succinate had been added in the presence of ATP, no reaction would have taken place, because of respiratory control.
(d) The mutation has little effect on the ability of the enzyme to catalyze the hydrolysis of ATP.
(e) They suggest two things: (1) the mutation did not affect the catalytic site on the enzyme, because ATP synthase is still capable of catalyzing the reverse reaction, and (2) the mutation did not affect the amount of enzyme present, given that the controls and patients had similar amounts of activity.

51. The absolute configuration of thiophosphate indicates that inversion at phosphorus has taken place in the reaction catalyzed by ATP synthase. This result is consistent with an inline phosphoryl-transfer reaction taking place in a single step. The retention of configuration in the Ca^{2+}-ATPase reaction points to two phosphoryl-transfer reactions—inversion by the first and a return to the starting configuration by the second. The Ca^{2+}-ATPase reaction proceeds by a phosphorylated enzyme intermediate.

Chapter 19

1. Ultimately, all of the carbon atoms of which we are made, not just carbohydrates, enter the biosphere through the process of photosynthesis. Moreover, the oxygen that we require is produced by photosynthesis.

2. $2\ NADP^+ + 3\ ADP^{3-} + 3\ P_i^{2-} + H^+ \rightarrow O_2 + 2\ NADPH + 3\ ATP^{4-} + H_2O$

3. (a) 7; (b) 5; (c) 4; (d) 10; (e) 1; (f) 2; (g) 9; (h) 3; (i) 8; (j) 6.

4. Photosystem II, in conjunction with the oxygen-generating complex, powers oxygen release. The reaction center of photosystem II absorbs light maximally at 680 nm.

5. Oxygen consumption will be maximal when photosystems I and II are operating cooperatively. Oxygen will be efficiently generated when electrons from photosystem II fill the electron holes in photosystem I, which were generated when the reaction center of photosystem I was illuminated by light of 700 nm.

6. Photosystem I generates ferredoxin, which reduces $NADP^+$ to NADPH, a biosynthetic reducing power. Photosystem II activates the water-oxidizing complex, generating electrons for photosynthesis, and generating protons to form a proton gradient and to reduce ferredoxin and O_2.

7. The light reactions take place on thylakoid membranes. Increasing the membrane surface increases the number of ATP- and NADPH-generating sites.

8. These complexes absorb more light than can a reaction center alone. The light-harvesting complexes funnel light to the reaction centers.

9. $NADP^+$ is the acceptor. H_2O is the donor. Light energy.

10. The charge gradient, a component of the proton-motive force in mitochondria, is neutralized by the influx of Mg^{2+} into the lumen of the thylakoid membranes.

11. Chlorophyll is readily inserted into the hydrophobic interior of the thylakoid membranes.

12. Protons released by the oxidation of water; protons pumped into the lumen by the cytochrome *bf* complex; protons removed from the stroma by the reduction of $NADP^+$ and plastoquinone.

13. 700-nm photons have an energy content of $+172$ kJ mol^{-1}. The absorption of light by photosystem I results in a $\Delta E_0'$ of -1.0 V. Recall that $\Delta G_0' = -nF\Delta E_0'$, where $F = 96.48$ kJ mol^{-1} V^{-1}. Under standard conditions, the energy change for the electrons is 96.5 kJ. Thus, the efficiency is $96.5/172 = 56\%$.

14. The electron flow from photosystem II to photosystem I is uphill, or exergonic. For this uphill flow, ATP would need to be consumed, defeating the purpose of photosynthesis.

15. $\Delta E_0' = 10.11$ V, and $\Delta G^{\circ\prime} = -21.3$ kJ mol^{-1} (-5.1 kcal mol^{-1}).

16. (a) All ecosystems require an energy source from outside the system, because the chemical energy sources will ultimately be limited. The photosynthetic conversion of sunlight is one example of such a conversion.
(b) Not at all. Spock would point out that chemicals other than water can donate electrons and protons.

17. DCMU inhibits electron transfer in the link between photosystems II and I. O_2 can evolve in the presence of DCMU if an artificial electron acceptor such as ferricyanide can accept electrons from Q.

18. DCMU will have no effect, because it blocks photosystem II, and cyclic photophosphorylation uses photosystem I and the cytochrome *bf* complex.

19. (a) $+120$ kJ $einstein^{-1}$ ($+28.7$ kcal $einstein^{-1}$)
(b) 1.24 V
(c) One 1000-nm photon has the free energy content of 2.4 molecules of ATP. A minimum of 0.42 photon is needed to drive the synthesis of a molecule of ATP.

20. At this distance, the expected rate is one electron per second.

21. The distance doubles, and so the rate should decrease by a factor of 64 to 640 ps.

22. The cristae.

23. In eukaryotes, both processes take place in specialized organelles. Both depend on high-energy electrons to generate ATP. In oxidative phosphorylation, the high-energy electrons originate in fuels and are extracted as reducing power in the form of NADH. In photosynthesis, the high-energy electrons are generated by light and are captured as reducing power in the form of NADPH. Both processes use redox reactions to generate a proton gradient, and the enzymes that convert the proton gradient into ATP are very similar in both processes. In both systems, electron transport takes place in membranes inside organelles.

24. We need to factor in the NADPH because it is an energy-rich molecule. Recall from Chapter 18 that NADH is worth 2.5 ATP if oxidized by the electron-transport chain. 12 NADPH = 30 ATP. Eighteen molecules of ATP are used directly, and so the equivalent of 48 molecules of ATP is required for the synthesis of glucose.

25. Both photosynthesis and cellular respiration are powered by high-energy electrons flowing toward a more-stable state. In cellular respiration, the high-energy electrons are derived from the oxidation of carbon fuels as NADH and $FADH_2$. They release their energy as they reduce oxygen. In photosynthesis, high-energy electrons are generated by absorbing light energy, and they find stability in photosystem I and ferridoxin.

26. The electrons flow through photosystem II directly to ferricyanide. No other steps are required.

27. (a) Thioredoxin

(b) The control enzyme is unaffected, but the mitochondrial enzyme with part of the chloroplast γ subunit increases activity as the concentration of DTT increases.

(c) The increase was even larger when thioredoxin was present. Thioredoxin is the natural reductant for the chloroplast enzyme, and so it presumably operates more efficiently than would DTT, which probably functions to keep the thioredoxin reduced.

(d) They seem to have done so.

(e) The enzyme is susceptible to control by the redox state. In plant cells, reduced thioredoxin is generated by photosystem I. Thus, the enzyme is active when photosynthesis is taking place.

(f) Cysteine.

(g) Group-specific modification or site-specific mutagenesis.

Chapter 20

1. The Calvin cycle is the primary means of converting gaseous CO_2 into organic matter—that is, biomolecules. Essentially, every carbon atom in your body passed through rubisco and the Calvin cycle at some time in the past.

2. Autotrophs can use the energy of sunlight, carbon dioxide, and water to synthesize carbohydrates, which can subsequently be used for catabolic or anabolic purposes. Heterotrophs require chemical fuels and are thus ultimately dependent on autotrophs.

3. Nothing grim or secret about these reactions. They are sometimes called the dark reactions because they do not directly depend on light.

4.

Calvin cycle	**Krebs cycle**
Stroma	Matrix
Carbon chemistry for photosynthesis	Carbon chemistry for oxidative phosphorylation
Fixes CO_2	Releases CO_2
Requires high-energy electrons (NADPH)	Generates high-energy electrons (NADH)
Regenerates starting compound (ribulose 1.5-bisphosphate)	Regenerates starting compound (oxaloacetate)
Requires ATP	Generates ATP
Complex stoichiometry	Simple stoichiometry

5. (a) 3-Phosphoglycerate. (b) The other members of the Calvin cycle.

6. Stage 1 is the fixation of CO_2 with ribulose 1,5-bisphosphate and the subsequent formation of 3-phosphoglycerate. Stage 2 is the conversion of some of the 3-phosphoglycerate into hexose. Stage 3 is the regeneration of ribulose 1,5-bisphosphate.

7. It catalyzes a crucial reaction, but it is highly inefficient. Consequently, it is required in large amounts to overcome its slow catalysis.

8. Because carbamate forms only in the presence of CO_2, this property prevents rubisco from catalyzing the oxygenase reaction exclusively when CO_2 is absent.

9. Because NADPH is generated in the chloroplasts by the light reactions.

10. The concentration of 3-phosphoglycerate would increase, whereas that of ribulose 1,5-bisphosphate would decrease.

11. The concentration of 3-phosphoglycerate would decrease, whereas that of ribulose 1,5-bisphosphate would increase.

12. Aspartate + glyoxylate $\longrightarrow$ oxaloacetate + glycine

13. The oxygenase activity of rubisco increases with temperature. Crabgrass is a C_4 plant, whereas most grasses lack this capability. Consequently, the crabgrass will thrive at the hottest part of the summer because the C_4 pathway provides an ample supply of CO_2.

14. The C_4 pathway allows the CO_2 concentration to increase at the site of carbon fixation. High concentrations of CO_2 inhibit the oxygenase reaction of rubisco. This inhibition is important for tropical plants because the oxygenase activity increases more rapidly with temperature than does the carboxylase activity.

15. ATP is required to form phosphoenolpyruvate (PEP) from pyruvate. The PEP combines with CO_2 to form oxaloacetate and, subsequently, malate. Two ATP molecules are required because a second ATP molecule is required to phosphorylate AMP to ADP.

16. Photorespiration is the consumption of oxygen by plants with the production of CO_2, but it does not generate energy. Photorespiration is due to the oxygenase activity of rubisco. It is wasteful because, instead of fixing CO_2 for conversion into hexoses, rubisco is generating CO_2.

17. As global warming progresses, C_4 plants will invade the higher latitudes, and C_3 plants will retreat to cooler regions.

18. The light reactions lead to an increase in the stromal concentrations of NADPH, reduced ferredoxin, and Mg^{2+}, as well as an increase in pH.

19. (a) 5; (b) 1; (c) 7; (d) 2; (e) 10; (f) 3; (g) 6; (h) 4; (i) 8; (j) 9.

20. The enzymes catalyze the transformation of the five-carbon sugar formed by the oxidative phase of the pentose phosphate pathway into fructose 6-phosphate and glyceraldehyde 3-phosphate, intermediates in glycolysis (and gluconeogenesis).

21. (a) C; (b) B and F; (c) G; (d) F; (e) E; (f) H; (g) I; (h) D; (i) A; (j) F; (k) B.

22. The label emerges at C-5 of ribulose 5-phosphate.

23. Oxidative decarboxylation of isocitrate to α-ketoglutarate. A β-ketoacid intermediate is formed in both reactions.

24. (a) 5 Glucose 6-phosphate + ATP $\rightarrow$ 6 ribose 5-phosphate + ADP + H^+.

(b) Glucose 6-phosphate + 12 $NADP^+$ + $7H_2O$ $\rightarrow$ 6 CO_2 + 12 NADPH + 12 H^+ + P_i.

25. The nonoxidative phase of the pentose phosphate pathway can be used to convert three molecules of ribose 5-phosphate into two molecules of fructose 6-phosphate and one molecule of glyceraldehyde 3-phosphate. These molecules are components of the glycolytic pathway.

26. The conversion of fructose 6-phosphate into fructose 1,6-bisphosphate by phosphofructokinase requires ATP.

27. When much NADPH is required. The oxidative phase of the pentose phosphate pathway is followed by the nonoxidative phase. The resulting fructose 6-phosphate and glyceraldehyde 3-phosphate are used to generate glucose 6-phosphate through gluconeogenesis, and the cycle is repeated until the equivalent of one glucose molecule is oxidized to CO_2.

28. Fava beans contain vicine, a purine glycoside that can lead to the generation of peroxides—reactive oxygen species that can damage membranes as well as other biomolecules. Glutathione is used to detoxify the ROS. The regeneration of glutathione depends on an adequate supply of NADPH, which is synthesized by the oxidative phase of the pentose phosphate pathway. People with low levels of the dehydrogenase are especially susceptible to vicine toxicity.

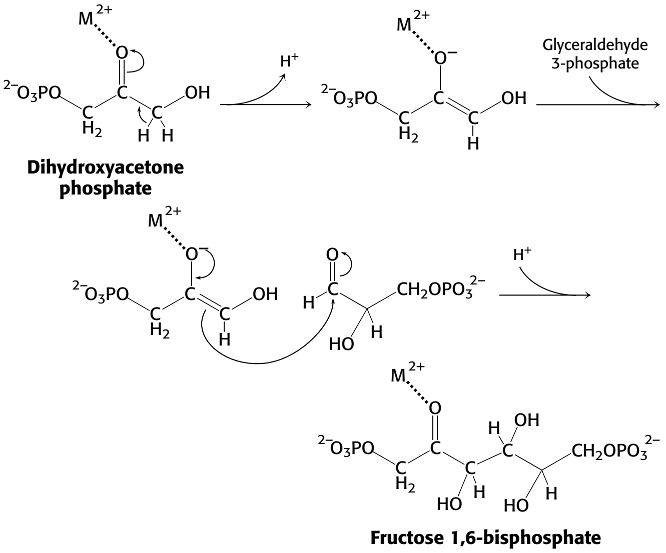
29. Because red blood cells do not have mitochondria and the only means to obtain NADPH is through the pentose phosphate pathway. There are biochemical means to convert mitochondrial NADH into cytoplasmic NADPH.

30. Reactive peroxides are a type of reactive oxygen species. The enzyme glutathione peroxidase uses reduced glutathione to neutralize peroxides by converting them into alcohols while generating

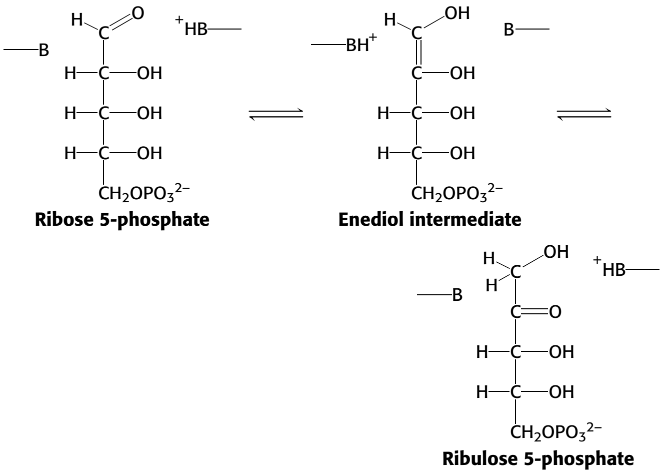
oxidized glutathione. Reduced glutathione is regenerated by glutathione reductase with the use of NADPH, the product of the oxidative phase of the pentose phosphate pathway.

31. $\Delta E'_0$ for the reduction of glutathione by NADPH is $+0.09$ V. Hence, $\Delta G^{\circ\prime}$ is -17.4 kJ mol^{-1} (-4.2 kcal mol^{-1}), which corresponds to an equilibrium constant of 1126. The required [NADPH]/[NADP$^+$] ratio is 8.9×10^{-5}.

32.


M^{2+}
$^{2-}O_3PO$
H^+
Glyceraldehyde 3-phosphate
OH
$CH_2OPO_3^{2-}$
Dihydroxyacetone phosphate
Fructose 1,6-bisphosphate

33.



34. Incubate an aliquot of a tissue homogenate with glucose labeled with ^{14}C at C-1, and incubate another with glucose labeled with ^{14}C at C-6. Compare the radioactivity of the CO_2 produced by the two samples. The rationale of this experiment is that only C-1 is decarboxylated by the pentose phosphate pathway, whereas C-1 and C-6 are decarboxylated equally when glucose is metabolized by the glycolytic pathway, the pyruvate dehydrogenase complex, and the citric acid cycle. The reason for the equivalence of C-1 and C-6 in the latter set of reactions is that glyceraldehyde 3-phosphate and dihydroxyacetone phosphate are rapidly interconverted by triose phosphate isomerase.

35. Lacking mitochondria, red blood cells metabolize glucose to lactate to obtain energy in the form of ATP. The CO_2 results from extensive use of the pentose phosphate pathway coupled with gluconeogenesis. This coupling allows the generation of much NADPH with the complete oxidation of glucose by the oxidative branch of the pentose phosphate pathway.

36. (a) $k_{\text{cat}}^{\text{CO}_2}/K_{\text{M}}^{\text{CO}_2} = 3 \times 10^6\ \text{s}^{-1}\text{M}^{-1}$

$k_{\text{cat}}^{\text{O}_2}/K_{\text{M}}^{\text{O}_2} = 4 \times 10^3\ \text{s}^{-1}\text{M}^{-1}$

(b) Despite the fact that the specificity constant for CO_2 as a substrate is much greater than that of O_2, the concentration of O_2 in the atmosphere is higher than that of CO_2, allowing the oxygenation reaction to occur.

37. The reduction of each mole of CO_2 to the level of a hexose requires two moles of NADPH. The reduction of NADP$^+$ is a two-electron process. Hence, the formation of two moles of NADPH requires the pumping of four moles of electrons by photosystem I. The electrons given up by photosystem I are replenished by photosystem II, which needs to absorb an equal number of photons. Hence, eight photons are needed to generate the required NADPH. The energy input of eight moles of photons is $+1594$ kJ ($+381$ kcal). Thus, the overall efficiency of photosynthesis under standard conditions is at least 477/1594, or 30%.

38. It is neither a violation nor a miracle. The equation on page 597 requires not only 18 ATP, but also 12 NADPH. These electrons, if transferred to NAD$^+$ and used in the electron-transport chain, would yield 30 ATP. Thus, the synthesis of glucose requires the equivalent of 48 ATP.

39. (a) The curve on the right in graph A was generated by the C_4 plant. Recall that the oxygenase activity of rubisco increases with temperature more rapidly than does the carboxylase activity. Consequently, at higher temperatures, the C_3 plants would fix less carbon. Because C_4 plants can maintain a higher CO_2 concentration, the rise in temperature is less deleterious.

(b) The oxygenase activity will predominate. Additionally, when the temperature rise is very high, the evaporation of water might become a problem. The higher temperatures can begin to damage protein structures as well.

(c) The C_4 pathway is a very effective active-transport system for concentrating CO_2, even when environmental concentrations are very low.

(d) With the assumption that the plants have approximately the same capability to fix CO_2, the C_4 pathway is apparently the rate-limiting step in C_4 plants.

Chapter 21

1. Step 1 is the release of glucose 1-phosphate from glycogen by glycogen phosphorylase. Step 2 is the formation of glucose 6-phosphate from glucose 1-phosphate, a reaction catalyzed by phosphoglucomutase. Step 3 is the remodeling of the glycogen by the transferase and the glucosidase.

2. (a) 8; (b) 3; (c) 6; (d) 5; (e) 9; (f) 2; (g) 10; (h) 1; (i) 4; (j) 7.

3. Glycogen is an important fuel reserve for several reasons. The controlled breakdown of glycogen and release of glucose increase the amount of glucose that is available between meals. Hence, glycogen serves as a buffer to maintain blood-glucose levels. Glycogen's role in maintaining blood-glucose levels is especially important because glucose is virtually the only fuel used by the brain, except during prolonged starvation. Moreover, the glucose from glycogen is readily mobilized and is therefore a good source of energy for sudden, strenuous activity. Unlike fatty acids, the released glucose can provide energy in the absence of oxygen and can thus supply energy for anaerobic activity.

4. As an unbranched polymer, α-amylose has only one nonreducing end. Therefore, only one glycogen phosphorylase molecule could degrade each α-amylose molecule. Because glycogen is highly branched, there are many nonreducing ends per molecule.

Consequently, many phosphorylase molecules can release many glucose molecules per glycogen molecule.

5. The patient has a deficiency of the branching enzyme.

6. In muscle, the *b* form of phosphorylase is activated by AMP. In the liver, the *a* form is inhibited by glucose. The difference corresponds to the difference in the metabolic role of glycogen in each tissue. Muscle uses glycogen as a fuel for contraction, whereas the liver uses glycogen to maintain blood-glucose levels.

7. Cells maintain the $[\text{P}_i]/[\text{glucose 1-phosphate}]$ ratio at greater than 100, substantially favoring phosphorolysis. We see here an example of how the cell can alter the free-energy change to favor a reaction taking place by altering the ratio of substrate and product.

8. The high level of glucose 6-phosphate in von Gierke disease, resulting from the absence of glucose 6-phosphatase or the transporter, shifts the allosteric equilibrium of phosphorylated glycogen synthase toward the active form.

9. The phosphoryl donor is glucose 1,6-bisphosphate, which is formed from glucose 1-phosphate and ATP in a reaction catalyzed by phosphoglucokinase.

10. The different manifestations correspond to the different roles of the liver and muscle. Liver glycogen phosphorylase plays a crucial role in the maintenance of blood-glucose levels. Recall that glucose is the primary fuel for the brain. Muscle glycogen phosphorylase provides glucose only for the muscle and, even then, only when the energy needs of the muscle are high, as during exercise. The fact that there are two different diseases suggests that there are two different isozymic forms of the glycogen phosphorylase—a liver-specific isozyme and a muscle-specific isozyme.

11. Water is excluded from the active site to prevent hydrolysis. The entry of water could lead to the formation of glucose rather than glucose 1-phosphate. A site-specific mutagenesis experiment is revealing in this regard. In phosphorylase, Tyr 573 is hydrogen bonded to the 2′-OH group of a glucose residue. The ratio of glucose 1-phosphate to glucose product is 9000 : 1 for the wild-type enzyme, and 500 : 1 for the Phe 573 mutant. Model building suggests that a water molecule occupies the site normally filled by the phenolic OH group of tyrosine and occasionally attacks the oxocarbonium ion intermediate to form glucose.

12. The amylase activity was necessary to remove all of the glycogen from the glycogenin. Recall that glycogenin synthesizes oligosaccharides of about eight glucose units, and then activity stops. Consequently, if the glucose residues are not removed by extensive amylase treatment, glycogenin will not function.

13. The substrate can be handed directly from the transferase site to the debranching site.

14. During exercise, [ATP] falls and [AMP] rises. Recall that AMP is an allosteric activator of glycogen phosphorylase *b*. Thus, even in the absence of covalent modification by phosphorylase kinase, glycogen is degraded.

15. Although glucose 1-phosphate is the actual product of the phosphorylase reaction, glucose 6-phosphate is a more versatile molecule with respect to metabolism. Among other fates, glucose-6-phosphate can be processed to yield energy or building blocks. In the liver, glucose 6-phosphate can be converted into glucose and released into the blood.

16. Epinephrine binds to its G-protein-coupled receptor. The resulting structural changes activate a G_α protein, which in turn activates adenylate cyclase. Adenylate cyclase synthesizes cAMP, which activates protein kinase A. Protein kinase A partly activates phosphoryl kinase, which phosphorylates and activates glycogen phosphorylase. The calcium released during muscle contraction also activates the phosphorylase kinase, leading to further stimulation of glycogen phosphorylase.

17. First, the signal-transduction pathway is shut down when the initiating hormone is no longer present. Second, the inherent GTPase activity of the G protein converts the bound GTP into inactive GDP. Third, phosphodiesterases convert cyclic AMP into AMP. Fourth, PP1 removes the phosphoryl group from glycogen phosphorylase, converting the enzyme into the usually inactive *b* form.

18. It prevents both from operating simultaneously, which would lead to a useless expenditure of energy.

19. All these symptoms suggest central nervous system problems. If exercise is exhaustive enough or the athlete has not prepared well enough or both, liver glycogen also can be depleted. The brain depends on glucose derived from liver glycogen. The symptoms suggest that the brain is not getting enough fuel (hypoglycemia).

20. Liver phosphorylase *a* is inhibited by glucose, which facilitates the R → T transition. This transition releases PP1, which inactivates glycogen breakdown and stimulates glycogen synthesis. Muscle phosphorylase is insensitive to glucose.

21. (a) 4; (b) 1; (c) 5; (d) 10; (e) 7; (f) 2; (g) 8; (h) 9; (i) 6; (j) 3.

22. Phosphoglucomutase, UDP-glucose pyrophosphorylase, pyrophosphatase, glycogenin, glycogen synthase, and branching enzyme.

23. The enzyme pyrophosphatase converts the pyrophosphate into two molecules of inorganic phosphate. This conversion renders the overall reaction irreversible.

$$\begin{array}{c} \text{Glucose 1-phosphate} + \text{UTP} \rightleftharpoons \text{UDP-glucose} + \text{PP}_\text{i} \\ \text{PP}_\text{i} + \text{H}_2\text{O} \longrightarrow 2\text{P}_\text{i} \\ \hline \text{Glucose 1-phosphate} + \text{UTP} \longrightarrow \text{UDP-glucose} + 2\text{Pi} \end{array}$$

24. The presence of high concentrations of glucose 6-phosphate indicates that glucose is abundant and that it is not being used by glycolysis. Therefore, this valuable resource is saved by incorporation into glycogen.

25. Free glucose must be phosphorylated at the expense of a molecule of ATP. Glucose 6-phosphate derived from glycogen is formed by phosphorolytic cleavage, sparing one molecule of ATP. Thus, the net yield of ATP when glycogen-derived glucose is processed to pyruvate is three molecules of ATP compared with two molecules of ATP from free glucose.

26. Breakdown: Phosphoglucomutase converts glucose 1-phosphate, liberated from glycogen breakdown, into glucose 6-phosphate, which can be released as free glucose (liver) or processed in glycolysis (muscle and liver). Synthesis: Converts glucose 6-phosphate into glucose 1-phosphate, which reacts with UTP to form UDP-glucose, the substrate for glycogen synthase.

27. $\text{Glycogen}_n + \text{P}_\text{i} \longrightarrow \text{glycogen}_{n-1} + \text{glucose 6-phosphate}$

$$\text{Glucose 6-phosphate} \longrightarrow \text{glucose 1-phosphate}$$

$$\text{UTP} + \text{glucose 1-phosphate} \longrightarrow \text{UDP-glucose} + 2\ \text{P}_\text{i}$$

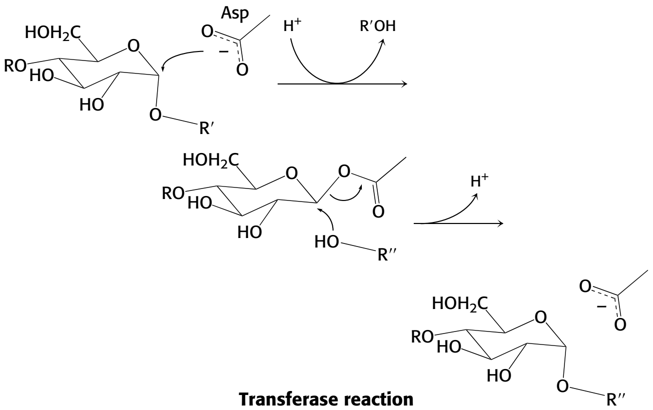
$$\text{Glycogen}_{n-1} + \text{UDP-glucose} \longrightarrow \text{glycogen}_n + \text{UDP}$$

$$\text{Sum: Glycogen}_n + \text{UTP} \longrightarrow \text{glycogen}_n + \text{UDP} + \text{P}_\text{i}$$

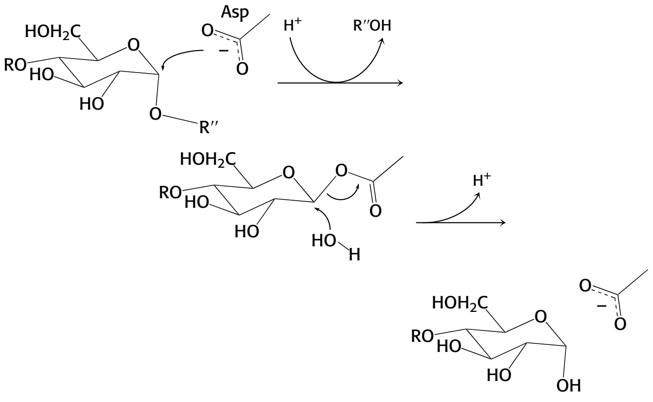
28. In principle, having glycogen be the only primer for the further synthesis of glycogen should be a successful strategy. However, if the glycogen granules were not evenly divided between daughter cells, glycogen stores for future generations of cells might be compromised. Glycogenin synthesizes the primer for glycogen synthase.

29. Insulin binds to its receptor and activates the tyrosine kinase activity of the receptor, which in turn triggers a pathway that activates protein kinases. The kinases phosphorylate and inactivate glycogen synthase kinase. Protein phosphatase 1 then removes the phosphate from glycogen synthase and thereby activates the synthase.

30.



Transferase reaction



α-1,6-Glucosidase reaction

31. In the liver, glucagon stimulates the cAMP-dependent pathway that activates protein kinase A. Epinephrine binds to a 7TM α-adrenergic receptor in the liver plasma membrane, which activates phospholipase C and the phosphoinositide cascade. This activation causes calcium ions to be released from the endoplasmic reticulum, which bind to calmodulin, and further stimulates phosphorylase kinase and glycogen breakdown.

32. Galactose + ATP + UTP + H_2O + $\text{glycogen}_n \longrightarrow$

glycogen_{n+1} + ADP + UDP + 2 P_i + H^+.

33. Phosphorylase, transferase, glucosidase, phosphoglucomutase, and glucose 6-phosphatase.

34. Glucose is an allosteric inhibitor of phosphorylase *a*. Hence, crystals grown in its presence are in the T state. The addition of glucose 1-phosphate, a substrate, shifts the R-to-T equilibrium toward the R state. The conformational differences between these states are sufficiently large that the crystal shatters unless it is stabilized by chemical cross-links.

35. Galactose is converted into UDP-galactose to eventually form glucose 6-phosphate.

36. This disease can also be produced by a mutation in the gene that encodes the glucose 6-phosphate transporter. Recall that glucose 6-phosphate must be transported into the lumen of the endoplasmic reticulum to be hydrolyzed by phosphatase. Mutations in the other three essential proteins of this system can likewise lead to von Gierke disease.

37. (a) Apparently, the glutamate, with its negatively charged R group, can mimic to some extent the presence of a phosphoryl group on serine. That the stimulation is not as great is not surprising in that the carboxyl group is smaller and not as charged as the phosphate.
(b) Substitution of aspartate would give some stimulation, but because it is smaller than the glutamate, the stimulation would be smaller.

38. (a) Glycogen was too large to enter the gel and, because analysis was by western blot with the use of an antibody specific to glycogenin, we would not expect to see background proteins.
(b) α-Amylase degrades glycogen, releasing the protein glycogenin, which can be visualized by a western blot.
(c) Glycogen phosphorylase, glycogen synthase, and protein phosphatase 1. These proteins might be visible if the gel were stained for protein, but a western analysis reveals the presence of glycogenin only.

39. (a) The smear was due to molecules of glycogenin with increasingly large amounts of glycogen attached to them.
(b) In the absence of glucose in the medium, glycogen is metabolized, resulting in a loss of the high-molecular-weight material.
(c) Glycogen could have been resynthesized and added to the glycogenin when the cells were fed glucose again.
(d) No difference between lanes 3 and 4 suggests that, by 1 hour, the glycogen molecules had attained maximum size in this cell line. Prolonged incubation does not apparently increase the amount of glycogen.
(e) α-Amylase removes essentially all of the glycogen, and so only the glycogenin remains.

Chapter 22

1. (a) 5; (b) 11; (c) 1; (d) 10; (e) 2; (f) 6; (g) 9; (h) 3; (i) 4; (j) 7; (k) 8.

2. Glycerol + 2 NAD^+ + P_i + ADP $\longrightarrow$

pyruvate + ATP + H_2O + 2 NADH + H^+

Glycerol kinase and glycerol phosphate dehydrogenase

3. The ready reversibility is due to the high-energy nature of the thioester in the acyl CoA.

4. To return the AMP to a form that can be phosphorylated by oxidative phosphorylation or substrate-level phosphorylation, another molecule of ATP must be expended in the reaction:

$$\text{ATP} + \text{AMP} \rightleftharpoons 2\ \text{ADP}$$

5. b, c, a, g, h, d, e, f.

6. The citric acid cycle. The reactions that take succinate to oxaloacetate, or the reverse, are similar to those of fatty acid metabolism (Section 17.2).

7. The next-to-last degradation product, acetoacetyl CoA, yields two molecules of acetyl CoA with the thiolysis by only one molecule of CoA.

8. Palmitic acid yields 106 molecules of ATP. Palmitoleic acid has a double bond between carbons C-9 and C-10. When palmitoleic acid is processed in β oxidation, one of the oxidation steps (to introduce a double bond before the addition of water) will not take place, because a double bond already exists. Thus, $FADH_2$ will not be generated, and palmitoleic acid will yield 1.5 fewer molecules of ATP than palmitic acid, for a total of 104.5 molecules of ATP.

9.

Activation fee to form the acyl CoA	−2 ATP
Seven rounds of yield:	
7 acetyl CoA at 10 ATP/acetyl CoA	+ 70 ATP
7 NADH at 2.5 ATP/NADH	+ 17.5 ATP
7 $FADH_2$ at 1.5 ATP/$FADH_2$	+ 10.5 ATP
Propionyl CoA, which requires an ATP to be converted into succinyl CoA	− 1 ATP
Succinyl CoA ⟶ succinate	+ 1 ATP
Succinate ⟶ fumarate + $FADH_2$ $FADH_2$ at 1.5 ATP/$FADH_2$	+ 1.5 ATP
Fumarate ⟶ malate	
Malate ⟶ oxaloacetate + NADH NADH at 2.5 ATP/NADH	+ 2.5 ATP
Total	120 ATP

10. You might hate yourself in the morning, but at least you won't have to worry about energy. To form stearoyl CoA requires the equivalent of 2 molecules of ATP.

$$\text{Stearoyl CoA} + 8\ \text{FAD} + 8\ \text{NAD}^+ + 8\ \text{CoA} + 8\ H_2O \longrightarrow 9\ \text{acetyl CoA} + 8\ \text{FADH}_2 + 8\ \text{NADH} + 8\ H^+$$

9 acetyl CoA at 10 ATP/acetyl CoA	+ 90 ATP
8 NADH at 2.5 ATP/NADH	+ 20 ATP
8 FADH_2 at 1.5 ATP/FADH_2	+ 12 ATP
Activation fee	−2.0
Total	122 ATP

11. Keep in mind that, in the citric acid cycle, 1 molecule of FADH_2 yields 1.5 ATP, 1 molecule of NADH yields 2.5 ATP, and 1 molecule of acetyl CoA yields 10 ATP. Two molecules of ATP are produced when glucose is degraded to 2 molecules of pyruvate. Two molecules of NADH also are produced, but the electrons are transferred to FADH_2 to enter the electron transport chain. Each molecule of FADH_2 can generate 1.5 ATP. Each molecule of pyruvate will produce 1 molecule of NADH. Each molecule of acetyl CoA generates 3 molecules of NADH, 1 molecule of FADH_2, and 1 molecule of ATP. So, we have a total of 10 ATP per acetyl CoA, or 20 for the 2 molecules of acetyl CoA. The total for glucose is 30 ATP. Now, what about hexanoic acid? Caprioic acid is activated to caprioic CoA at the expense of 2 ATP, and so we are 2 ATP in the hole. The first cycle of β oxidation generates 1 FADH_2, 1 NADH, and 1 acetyl CoA. After the acetyl CoA has been run through the citric acid cycle, this step will have generated a total of 14 ATP. The second cycle of β oxidation generates 1 FADH_2 and 1 NADH but 2 acetyl CoA. After the acetyl CoA has been run through the citric acid cycle, this step will have generated a total of 24 ATP. The total is 36 ATP. Thus, the foul-smelling caprioic acid has a net yield of 36 ATP. So on a per carbon basis, this fat yields 20% more ATP than does glucose, a manifestation of the fact that fats are more reduced than carbohydrates.

12. $\text{Stearate} + \text{ATP} + 13.5\ H_2O + 8\ \text{FAD} + 8\ \text{NAD}^+ \rightarrow 4.5\ \text{acetoacetate} + 14.5\ H^+ + 8\ \text{FADH}_2 + 8\ \text{NADH} + \text{AMP} + 2\ P_i$.

13. Palmitate is activated and then processed by β oxidation according to the following reactions.

$$\text{Palmitate} + \text{CoA} + \text{ATP} \longrightarrow \text{palmitoyl CoA} + \text{AMP} + 2\ P_i$$

$$\text{Palmitoyl CoA} + 7\ \text{FAD} + 7\ \text{NAD}^+ + 7\ \text{CoASH} + H_2O \longrightarrow 8\ \text{acetyl CoA} + 7\ \text{FADH}_2 + 7\ \text{NADH} + 7\ H^+$$

The eight molecules of acetyl CoA combine to form four molecules of acetoacetate for release into the blood, and so they do not contribute to the energy yield in the liver. However, the FADH_2 and NADH generated in the preparation of acetyl CoA can be processed by oxidative phosphorylation to yield ATP.

$$1.5\ \text{ATP}/\text{FADH}_2 \times 7 = 10.5\ \text{ATP}$$

$$2.5\ \text{ATP}/\text{NADH} \times 7 = 17.5\ \text{ATP}$$

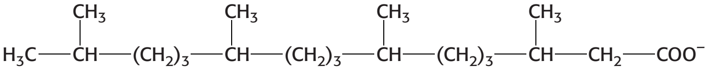
The equivalent of 2 ATP were used to form palmitoyl CoA. Thus, 26 ATP were generated for use by the liver.

14. NADH produced with the oxidation to acetoacetate = 2.5 ATP. Acetoacetate is converted into acetoacetyl CoA. Two molecules of acetyl CoA result from the hydrolysis of acetoacetyl CoA, each worth 10 ATP when processed by the citric acid cycle. Total ATP yield is 22.5.

15. Because a molecule of succinyl CoA is used to form acetoacetyl CoA. Succinyl CoA could be used to generate one molecule of ATP, and so someone could argue that the yield is 21.5.

16. For fats to be combusted, not only must they be converted into acetyl CoA, but the acetyl CoA must be processed by the citric acid cycle. In order for acetyl CoA to enter the citric acid cycle, there must be a supply of oxaloacetate. Oxaloacetate can be formed by the metabolism of glucose to pyruvate and the subsequent carboxylation of pyruvate to form oxaloacetate.

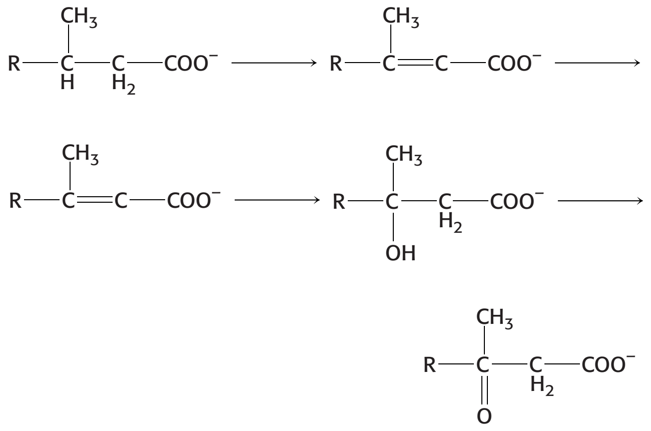
17. (a)



$$H_3C{-}\overset{CH_3}{\overset{|}{CH}}{-}(CH_2)_3{-}\overset{CH_3}{\overset{|}{CH}}{-}(CH_2)_3{-}\overset{CH_3}{\overset{|}{CH}}{-}(CH_2)_3{-}\overset{CH_3}{\overset{|}{CH}}{-}CH_2{-}COO^-$$

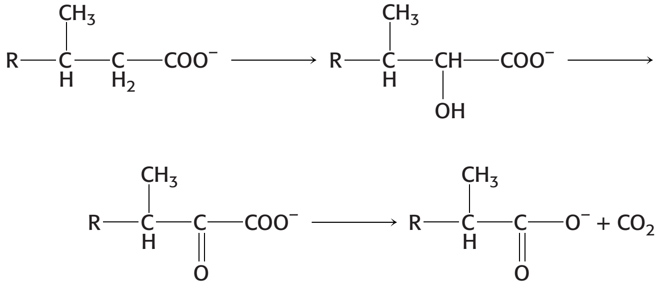
Phytanic acid

The problem with phytanic acid is that, as it undergoes β oxidation, we encounter the dreaded pentavalent carbon atom. Because the pentavalent carbon atom doesn't exist, β oxidation cannot take place and phytanic acid accumulates.



The dreaded pentavalent carbon atom

(b) Removing methyl groups, though theoretically possible, would be time consuming and lacking in elegance. What would we do with the methyl groups? Our livers solve the problem by inventing α oxidation.



One round of α oxidation rather than β oxidation converts phytanic acid into a β-oxidation substrate.

18. The first oxidation removes two tritium atoms. The hydration adds nonradioactive H and OH. The second oxidation removes another tritium atom from the β-carbon atom. Thiolysis removes an acetyl CoA with only one tritium atom; so the tritium-to-carbon ratio is 1/2. This ratio will be the same for two of the acetates. The last one, however, does not undergo oxidation, and so all tritium remains. The ratio for this acetate is 3/2. The ratio for the entire molecule is then 5/6.

19. In the absence of insulin, lipid mobilization will take place to an extent that it overwhelms the ability of the liver to convert the lipids into ketone bodies.

20. (a) 10; (b) 1; (c) 5; (d) 8; (e) 3; (f) 9; (g) 6; (h) 7; (i) 4; (j) 2.

21. (a) Oxidation in mitochondria; synthesis in the cytoplasm.
(b) Coenzyme A in oxidation; acyl carrier protein for synthesis.
(c) FAD and NAD^+ in oxidation; NADPH for synthesis.
(d) The L isomer of 3-hydroxyacyl CoA in oxidation; the D isomer in synthesis.
(e) From carboxyl to methyl in oxidation; from methyl to carboxyl in synthesis.
(f) The enzymes of fatty acid synthesis, but not those of oxidation, are organized in a multienzyme complex.

22. Bicarbonate is required for the synthesis of malonyl CoA from acetyl CoA by acetyl CoA carboxylase.

23. $$7 \text{ acetyl CoA} + 6 \text{ ATP} + 12 \text{ NADPH} + 12 \text{ H}^+ \longrightarrow$$
$$\text{myristate} + 7 \text{ CoA} + 6 \text{ ADP} + 6 \text{ P}_i + 12 \text{ NADP}^+ + 5\text{H}_2\text{O}.$$

24. We will need six acetyl CoA units. One acetyl CoA unit will be used directly to become the two carbon atoms farthest from the acid end. The other five units must be converted into malonyl CoA. The synthesis of each malonyl CoA molecule costs a molecule of ATP; so 5 molecules of ATP will be required. Each round of elongation requires 2 molecules of NADPH, 1 molecule to reduce the keto group to an alcohol and 1 molecule to reduce the double bond. As a result, 10 molecules of NADPH will be required. Therefore, 5 molecules of ATP and 10 molecules of NADPH are required to synthesize lauric acid.

25. e, b, d, a, c.

26. Such a mutation would inhibit fatty acid synthesis because the enzyme cleaves cytoplasmic citrate to yield acetyl CoA for fatty acid synthesis.

27. (a) False. Biotin is required for acetyl CoA carboxylase activity.
(b) True.
(c) False. ATP is required to synthesize malonyl CoA.
(d) True.
(e) True.
(f) False. Fatty acid synthase is a dimer.
(g) True.
(h) False. Acetyl CoA carboxylase is stimulated by citrate, which is cleaved to yield its substrate acetyl CoA.

28. Fatty acids with odd numbers of carbon atoms are synthesized starting with propionyl ACP (instead of acetyl ACP), which is formed from propionyl CoA by acetyl transacetylase.

29. All of the labeled carbon atoms will be retained. Because we need 8 acetyl CoA molecules and only 1 carbon atom is labeled in the acetyl group, we will have 8 labeled carbon atoms. The only acetyl CoA used directly will retain 3 tritium atoms. The 7 acetyl CoA molecules used to make malonyl CoA will lose 1 tritium atom on addition of the CO_2 and another one at the dehydration step. Each of the 7 malonyl CoA molecules will retain 1 tritium atom. Therefore, the total retained tritium is 10 atoms. The ratio of tritium to carbon is 1.25.

30. With a diet rich in raw eggs, avidin will inhibit fatty acid synthesis by reducing the amount of biotin required by acetyl CoA carboxylase. Cooking the eggs will denature avidin, and so it will no longer bind biotin.

31. The only acetyl CoA used directly, not in the form of malonyl CoA, provides the two carbon atoms at the ω end of the fatty acid chain. Because palmitic acid is a C_{16} fatty acid, acetyl CoA will have provided carbons 15 and 16.

32. HCO_3^- is attached to acetyl CoA to form malonyl CoA. When malonyl CoA condenses with acetyl CoA to form the four-carbon ketoacyl CoA, the HCO_3^- is lost as CO_2.

33. Phosphofructokinase controls the flux down the glycolytic pathway. Glycolysis functions to generate ATP or building blocks for biosynthesis, depending on the tissue. The presence of citrate in the cytoplasm indicates that those needs are met, and there is no need to metabolize glucose.

34. C-1 is more radioactive.

35. The mutant enzyme will be persistently active because it cannot be inhibited by phosphorylation. Fatty acid synthesis will be abnormally active. Such a mutation might lead to obesity.

36. (a) Palmitoleate; (b) linoleate; (c) linoleate; (d) oleate; (e) oleate; (f) linolenate.

37. Decarboxylation drives the condensation of malonyl ACP and acetyl ACP. In contrast, the condensation of two molecules of acetyl ACP is energetically unfavorable. In gluconeogenesis, decarboxylation drives the formation of phosphoenolpyruvate from oxaloacetate.

38. Fat mobilization in adipocytes is activated by phosphorylation. Hence, overproduction of the cAMP-activated kinase will lead to an accelerated breakdown of triacylglycerols and a depletion of fat stores.

39. Carnitine translocase deficiency and glucose 6-phosphate transporter deficiency.

40. In the fifth round of β oxidation, *cis*-Δ^2-enoyl CoA is formed. Dehydration by the classic hydratase yields D-3-hydroxyacyl CoA, the wrong isomer for the next enzyme in β oxidation. This dead end is circumvented by a second hydratase that removes water to give *trans*-Δ^2-enoyl CoA. The addition of water by the classic hydratase then yields L-3-hydroxyacyl CoA, the appropriate isomer. Thus, hydratases of opposite stereospecificities serve to epimerize (invert the configuration of) the 3-hydroxyl group of the acyl CoA intermediate.

41. An advantage of this arrangement is that the synthetic activity of different enzymes is coordinated. In addition, intermediates can be efficiently handed from one active site to another without leaving the assembly. Furthermore, a complex of covalently joined enzymes is more stable than one formed by noncovalent attractions. Each of the component enzymes is recognizably homologous to its bacterial counterpart.

42. The probability of synthesizing an error-free polypeptide chain decreases as the length of the chain increases. A single mistake can make the entire polypeptide ineffective. In contrast, a defective subunit can be spurned in the formation of a noncovalent multienzyme complex; the good subunits are not wasted.

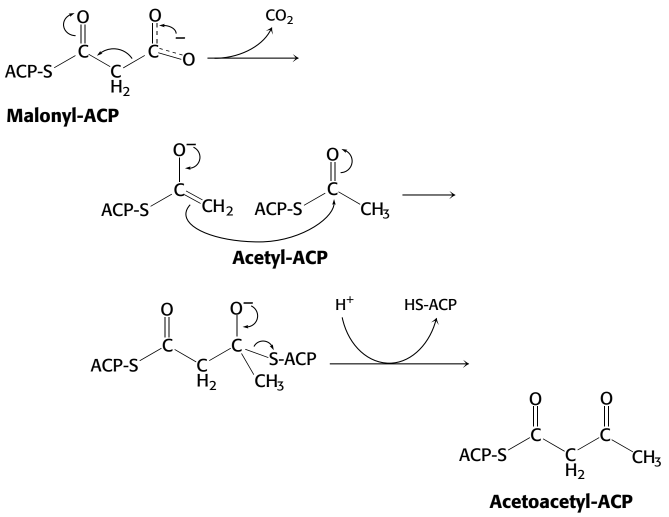
43. The absence of ketone bodies is due to the fact that the liver, the source of ketone bodies in the blood, cannot oxidize fatty acids to produce acetyl CoA. Moreover, because of the impaired fatty acid oxidation, the liver becomes more dependent on glucose as an energy source. This dependency results in a decrease in gluconeogenesis and a drop in blood-glucose levels, which is exacerbated by the lack of fatty acid oxidation in muscle and a subsequent increase in glucose uptake from the blood.

44. Peroxisomes enhance the degradation of very long chain fatty acids. Consequently, increasing the activity of peroxisomes could help to lower levels of blood triglycerides. In fact, clofibrate is rarely used because of serious side effects.

45. Citrate works by facilitating, in cooperation with the protein MIG12, the formation of active filaments from inactive monomers. In essence, it increases the number of active sites available, or the concentration of enzyme. Consequently, its effect is visible as an increase in the value of $V_{\max}$. Allosteric enzymes that alter their $V_{\max}$ values in response to regulators are sometimes called V-class enzymes. The more common type of allosteric enzyme, in which K_m is altered, comprises K-class enzymes. Palmitoyl CoA causes depolymerization and thus inactivation.

46. The thiolate anion of CoA attacks the 3-keto group to form a tetrahedral intermediate. This intermediate collapses to form acyl CoA and the enolate anion of acetyl CoA. Protonation of the enolate yields acetyl CoA.

47.



48. (a) Fats burn in the flame of carbohydrates. Without carbohydrates, there would be no anapleurotic reactions to replenish the components of the citric acid cycle. With a diet of fats only, the acetyl CoA from fatty acid degradation would build up.
(b) Acetone from ketone bodies.
(c) Yes. Odd-chain fatty acids would lead to the production of propionyl CoA, which can be converted into succinyl CoA, a citric acid cycle component. It would serve to replenish the citric acid cycle, allow glucose synthesis and mitigate the halitosis.

49. A labeled fat can enter the citric acid cycle as acetyl CoA and yield labeled oxaloacetate, but only after two carbon atoms have been lost as CO_2. Consequently, even though oxaloacetate may be labeled, there can be no net synthesis of oxaloacetate and hence no net synthesis of glucose or glycogen.

50. (a) Glucose is the primary fuel used by the brain. Lack of pyruvate dehydrogenase would prevent complete oxidation of glucose-derived pyruvate by cellular respiration. (b) Because ATP cannot be generated aerobically, glucose would be metabolized to lactate to obtain some ATP. (c) Such a diet would generate ketone bodies for use as fuel by the brain.

51. I-cell disease (Section 11.3) results when lysosomal enzymes are secreted instead of imported into the lysosome.

52. (a) The $V_{\max}$ is decreased and the K_m is increased. $V_{\max}$ (wild type) = 13 nmol minute^{-1} mg^{-1}; K_m (wild type) = 45 μM; $V_{\max}$ (mutant) = 8.3 nmol minute^{-1} mg^{-1}; K_m (mutant) = 74 μM.
(b) Both the $V_{\max}$ and the K_m are decreased. $V_{\max}$ (wild type) = 41 nmol minute^{-1} mg^{-1}; K_m (wild type) = 104 μM; $V_{\max}$ (mutant) = 23 nmol minute^{-1} mg^{-1}; K_m (mutant) = 69 μM.
(c) The wild type is significantly more sensitive to malonyl CoA.
(d) With respect to carnitine, the mutant displays approximately 65% of the activity of the wild type; with respect to palmitoyl CoA, approximately 50% activity. On the other hand, 10 μM of malonyl CoA inhibits approximately 80% of the wild type but has essentially no effect on the mutant enzyme.
(e) The glutamate appears to play a more prominent role in regulation by malonyl CoA than in catalysis.

53. (a) Leucine is changed to alanine.
(b) No change relative to control.
(c) Tyrosine to alanine.
(d) Greatly reduced lipase binding compared to the control.
(e) Lipase activity was diminished.
(f) This suggests that they are separate. The L16A mutation allows binding of the lipase but apparently does not allow activation of the lipase.
(g) This mutation decreased both binding and activity, suggesting that the tyrosine is involved in both aspects of colipase activity.

54. (a) Soraphen A inhibits fatty acid synthesis in a dose-dependent manner.
(b) Fatty acid oxidation is increased in the presence of soraphen A.
(c) Recall that acetyl CoA carboxylase 2 synthesizes malonyl CoA to inhibit the transport of fatty acids into the mitochondria, thereby preventing fatty acid oxidation. Soraphen A apparently inhibits both forms of the carboxylase.
(d) Phospholipid synthesis was inhibited in a dose-dependent manner.
(e) Phospholipids are required for membrane synthesis.
(f) Soraphen A inhibits cell proliferation, especially at higher concentrations.

Chapter 23

1. When the proteins are denatured, all of the peptide bonds are accessible to proteolytic enzymes. If the three-dimensional structure of a protein is maintained, access to many peptide bonds is denied to the proteolytic enzymes.

2. First, the ubiquitin-activating enzyme (E1) links ubiquitin to a sulfhydryl group on E1 itself. Next, the ubiquitin is transferred to a cysteine residue on the ubiquitin-conjugating enzyme (E2) by E2. The ubiquitin–protein ligase (E3), using the ubiquitinated E2 as a substrate, transfers the ubiquitin to the target protein.

3. (a) 7; (b) 4; (c) 2; (d) 10; (e) 5; (f) 3; (g) 9; (h) 1; (i) 6; (j) 8.

4. (a) The ATPase activity of the 26S proteasome resides in the 19S subunit. The energy of ATP hydrolysis is used to unfold the substrate, which is too large to enter the catalytic barrel. ATP may also be required for translocation of the substrate into the barrel.
(b) Substantiates the answer in part *a*. Because they are small, the peptides do not need to be unfolded. Moreover, small peptides could probably enter all at once and not require translocation.

5. (a) Pyruvate; (b) oxaloacetate; (c) α-ketoglutarate; (d) α-ketoisocaproate; (e) phenylpyruvate; (f) hydroxyphenylpyruvate.

6. (a) Aspartate + α-ketoglutarate + GTP + ATP + 2 H_2O + NADH + H^+ → ½ glucose + glutamate + CO_2 + ADP + GDP + NAD^+ + 2 P_i.
The required coenzymes are pyridoxal phosphate in the transamination reaction and NAD^+/NADH in the redox reactions.
(b) Aspartate + CO_2 + NH_4^+ + 3 ATP + NAD^+ + 4 H_2O → oxaloacetate + urea + 2 ADP + 4 P_i + AMP + NADH + H^+.

7. In the eukaryotic proteasome, the distinct β subunits have different substrate specificities, allowing proteins to be more thoroughly degraded.

8. The six subunits probably exist as a heterohexamer. Cross-linking experiments could test the model and help determine which subunits are adjacent to one another.

9. Thiamine pyrophosphate.

10. Aminotransferases transfer the α-amino group to α-ketoglutarate to form glutamate. Glutamate is oxidatively deaminated to form an ammonium ion.

11. Aspartate (oxaloacetate), glutamate (α-ketoglutarate), alanine (pyruvate).

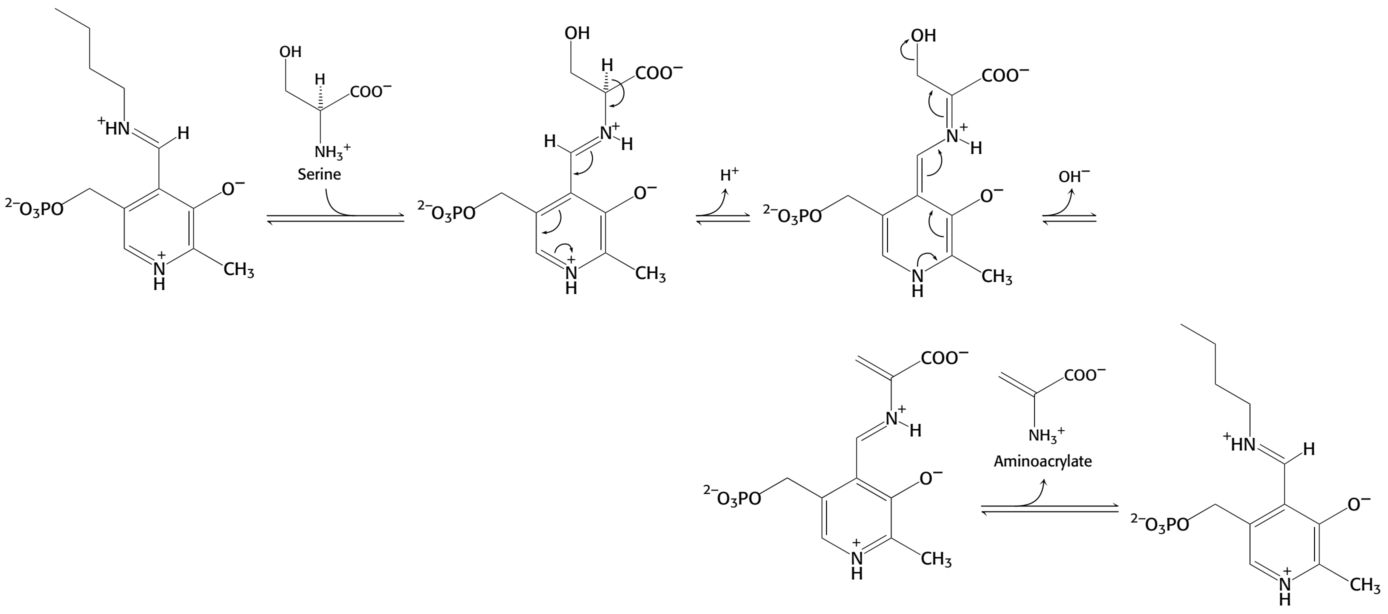
12. Serine and threonine.

13. They are either fuels for the citric acid cycle, components of the citric acid cycle, or molecules that can be converted into a fuel for the citric acid cycle in one step.

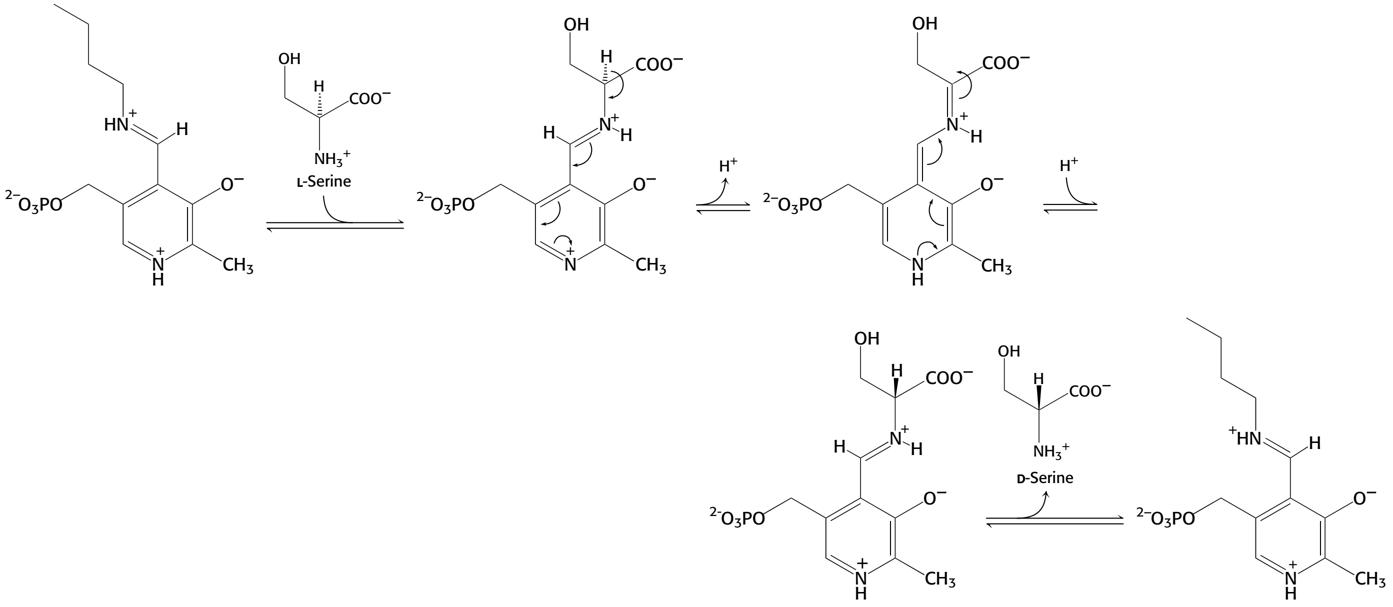
14. It acts as an electron sink.
15. Carbamoyl phosphate and aspartate.
16. (a) 4; (b) 5; (c) 1; (d) 6; (e) 7; (f) 3; (g) 2.
17. A, arginine; B, citrulline; C, ornithine; D, arginosuccinate. The order of appearance: C, B, D, A.
18. $$CO_2 + NH_4^+ + 3\,ATP + NAD^+ + \text{aspartate} + 3\,H_2O \longrightarrow \text{urea} + 2\,ADP + 2\,P_i + AMP + PP_i + NADH + H^+ + \text{oxaloacetate}.$$

Four high-transfer-potential phosphoryl groups are spent. Note, however, that an NADH is generated if fumarate is converted into oxaloacetate. NADH can generate 2.5 ATP in the electron-transport chain. Taking these ATP into account, only 1.5 high-transfer-potential phosphoryl groups are spent.
19. The synthesis of fumarate by the urea cycle is important because it links the urea cycle and the citric acid cycle. Fumarate is hydrated to malate, which, in turn, is oxidized to oxaloacetate. Oxaloacetate has several possible fates: (1) transamination to aspartate, (2) conversion into glucose by the gluconeogenic pathway, (3) condensation with acetyl CoA to form citrate, or (4) conversion into pyruvate. You can collect.
20. Ornithine transcarbamoylase (Compound A is analogous to PALA; Chapter 10).
21. Ammonia could lead to the amination of α-ketoglutarate, producing a high concentration of glutamate in an unregulated fashion. α-Ketoglutarate for glutamate synthesis could be removed from the citric acid cycle, thereby diminishing the cell's respiration capacity.
22. The mass spectrometric analysis strongly suggests that three enzymes—pyruvate dehydrogenase, α-ketoglutarate dehydrogenase, and the branched-chain α-ketoacid dehydrogenase—are deficient. Most likely, the common E_3 component of these enzymes is missing or defective. This proposal could be tested by purifying these three enzymes and assaying their ability to catalyze the regeneration of lipoamide.
23. Benzoate, phenylacetate, and arginine would be given to supplement a protein-restricted diet. Nitrogen would emerge in hippurate, phenylacetylglutamine, and citrulline.
24. The liver is the primary tissue for capturing nitrogen as urea. If the liver is damaged (for instance, by hepatitis or the excessive consumption of alcohol), free ammonia is released into the blood.
25. This defect can be partly bypassed by providing a surplus of arginine in the diet and restricting the total protein intake. In the liver, arginine is split into urea and ornithine, which then reacts with carbamoyl phosphate to form citrulline. This urea-cycle intermediate condenses with aspartate to yield argininosuccinate, which is then excreted. Note that two nitrogen atoms—one from carbamoyl phosphate and the other from aspartate—are eliminated from the body per molecule of arginine provided in the diet. In essence, argininosuccinate substitutes for urea in carrying nitrogen out of the body. The formation of argininosuccinate removes the nitrogen, and the restriction on protein intake relieves the aciduria.
26. Aspartame, a dipeptide ester (L-aspartyl- L-phenylalanine methyl ester), is hydrolyzed to L-aspartate and L-phenylalanine. High levels of phenylalanine are harmful in phenylketonurics.
27. *N*-Acetylglutamate is synthesized from acetyl CoA and glutamate. Once again, acetyl CoA serves as an activated acetyl donor. This reaction is catalyzed by *N*-acetylglutamate synthase.
28. Not all proteins are created equal: some are more important than others. Some proteins would be degraded to provide the missing amino acid. The nitrogen from the other amino acids would be excreted as urea. Consequently, more nitrogen would be excreted than ingested.
29. The carbon skeletons of ketogenic amino acids can be converted into ketone bodies or fatty acids. Only leucine and lysine are purely ketogenic. Glucogenic amino acids are those whose carbon skeletons can be converted into glucose.
30. As shown in Figure 23.28, alanine, a gluconeogenic amino acid, is released during the metabolism of tryptophan to acetyl CoA and acetoacetyl CoA.
31. The branched-chain amino acids leucine, isoleucine, and valine. The required enzyme is the branched-chain α-ketoacid dehydrogenase complex.
32. Pyruvate (glycolysis and gluconeogenesis), acetyl CoA (citric acid cycle and fatty acid synthesis), acetoacetyl CoA (ketone-body formation), α-ketoglutarate (citric acid cycle), succinyl CoA (citric acid cycle), fumarate (citric acid cycle), and oxaloacetate (citric acid cycle and gluconeogenesis).

33.



34.



The equilibrium constant for the interconversion of L-serine and D-serine is exactly 1.

35. Double-displacement. A substituted enzyme intermediate is formed.

36. Exposure of such a domain suggests that a component of a multiprotein complex has failed to form properly or that one component has been synthesized in excess. This exposure leads to rapid degradation and the restoration of appropriate stoichiometries.

37. (a) Depletion of glycogen stores. When they are gone, proteins must be degraded to meet the glucose needs of the brain. The resulting amino acids are deaminated, and the nitrogen atoms are excreted as urea.

(b) The brain has adapted to the use of ketone bodies, which are derived from fatty acid catabolism. In other words, the brain is being powered by fatty acid breakdown.

(c) When the glycogen and lipid stores are gone, the only available energy source is protein.

38. The precise cause of all of the symptoms is not firmly established, but a likely explanation depends on the centrality of oxaloacetate to metabolism. A lack of pyruvate carboxylase would reduce the amount of oxaloacetate. The lack of oxaloacetate would reduce the activity of the citric acid cycle and so ATP would be generated by lactic acid formation. If the concentration of oxaloacetate is low, aspartate cannot be formed and the urea cycle would be compromised. Oxaloacetate is also required to form citrate, which transports acetyl CoA to the cytoplasm for fatty acid synthesis. Finally, oxaloacetate is required for gluconeogenesis.

39. Deamination to α-keto-β-methylvalerate; oxidative decarboxylation to α-methylbutyryl CoA; oxidation to tiglyl CoA; hydration, oxidation, and thiolysis yield acetyl CoA and propionyl CoA; propionyl CoA to succinyl CoA.

40. Glycogen phosphorylase. The coenzyme serves as an acid–base catalyst.

41. In the Cori cycle, the carbon atoms are transferred from muscle to liver as lactate. For lactate to be of any use, it must be reduced to pyruvate. This reduction requires high-energy electrons in the form of NADH. When the carbon atoms are transferred as alanine, transamination yields pyruvate directly.

42. (a) Virtually no digestion in the absence of nucleotides.

(b) Protein digestion is greatly stimulated by the presence of ATP.

(c) AMP-PNP, a nonhydrolyzable analog of ATP, is no more effective than ADP.

(d) The proteasome requires neither ATP nor PAN to digest small substrates.

(e) PAN and ATP hydrolysis may be required to unfold the peptide and translocate it into the proteasome.

(f) Although *Thermoplasma* PAN is not as effective with the other proteasomes, it nonetheless results in threefold to fourfold stimulation of digestion.

(g) In light of the fact that the archaea and eukarya diverged several billion years ago, the fact that *Thermoplasma* PAN can stimulate rabbit muscle suggests homology not only between the proteasomes, but also between PAN and the 19S subunit (most likely the ATPases) of the mammalian 26S proteasome.

Chapter 24

1. Nitrogen fixation is the conversion of atmospheric N_2 into NH_4^+. Diazotrophic (nitrogen-fixing) microorganisms are able to fix nitrogen.

2. (a) 4; (b) 8; (c) 10; (d) 6; (e) 7; (f) 9; (g) 3; (h) 5; (i) 2; (j) 1.

3. The reductase provides electrons with high reducing power, whereas the nitrogenase, which requires ATP hydrolysis, uses the electrons to reduce N_2 to NH_3.

4. False. Nitrogen is thermodynamically favored. ATP expenditure by the nitrogenase is required to make the reaction kinetically possible.

5. The bacteria provide the plant with ammonia by reducing atmospheric nitrogen. This reduction is energetically expensive, and the bacteria use ATP from the plant.

6. Oxaloacetate, pyruvate, ribose-5-phosphate, phosphoenolpyruvate, erythrose-4-phosphate, α-ketoglutarate, and 3-phosphoglycerate.

7. Human beings do not have the biochemical pathways to synthesize certain amino acids from simpler precursors. Consequently, these amino acids are “essential” and must be obtained from the diet.

8. Glucose + 2 ADP + 2 $\mathrm{P_i}$ + 2 $\mathrm{NAD^+}$ + 2 glutamate $\longrightarrow$ 2 alanine + 2 α-ketoglutrate + 2 ATP + 2 NADH + 2 $\mathrm{H_2O}$ + 2 $\mathrm{H^+}$.

9. $\mathrm{N_2} \longrightarrow \mathrm{NH_4}^+ \longrightarrow$ glutamate $\longrightarrow$ serine $\longrightarrow$ glycine $\longrightarrow$ δ-aminolevulinate $\longrightarrow$ porphobilinogen $\longrightarrow$ heme.

10. Pyridoxal phosphate (PLP).

11. *S*-Adenosylmethionine, tetrahydrofolate, and methylcobalamin.

12. (a) N^5,N^{10}-Methylenetetrahydrofolate; (b) N^5-methyltetrahydrofolate.

13. γ-Glutamyl phosphate is a likely reaction intermediate.

14. The synthesis of asparagine from aspartate passes through an acyl-adenylate intermediate. One of the products of the reaction will be ^{18}O-labeled AMP.

15. The administration of glycine leads to the formation of isovalerylglycine. This water-soluble conjugate, in contrast with isovaleric acid, is excreted very rapidly by the kidneys.

16. The nitrogen atom shaded red is derived from glutamine. The carbon atom shaded blue is derived from serine.

17. They carry out nitrogen fixation. The absence of photosystem II provides an environment in which $\mathrm{O_2}$ is not produced. Recall that the nitrogenase is very rapidly inactivated by $\mathrm{O_2}$.

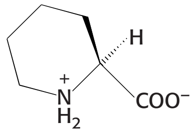
18. The cytoplasm is a reducing environment, whereas the extracellular milieu is an oxidizing environment.

19. (a) None; (b) D-glutamate and oxaloacetate.

20. Succinyl CoA is formed in the mitochondrial matrix as part of the citric acid cycle.

21. Alanine from pyruvate; aspartate from oxaloacetate; glutamate from α-ketoglutarate.

22. Lysine cyclodeaminase converts L-lysine into the six-membered ring analog of proline, also referred to as L-homoproline or L-pipecolate:



Pipecolate

23. Y could inhibit the C → D step, Z could inhibit the C → F step, and C could inhibit A → B. This scheme is an example of sequential feedback inhibition. Alternatively, Y could inhibit the C → D step, Z could inhibit the C → F step, and the A → B step would be inhibited only in the presence of both Y and Z. This scheme is called concerted feedback inhibition.

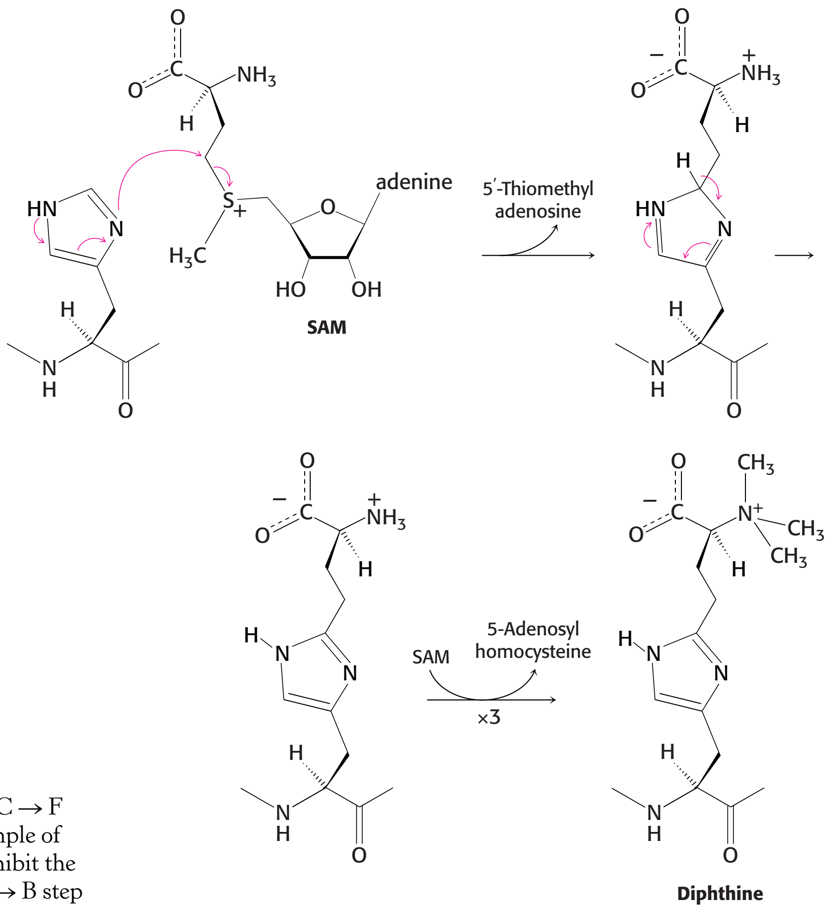
24. The rate of the A → B step in the presence of high levels of Y and Z would be 24 $\mathrm{s^{-1}}$ (0.6 × 0.4 × 100 $\mathrm{s^{-1}}$).

25. Lysine 258 is absolutely essential for the activity of aspartate aminotransferase, as it is responsible both for the formation of the internal aldimine with the pyridoxal phosphate cofactor and for transferring the proton between the ketimine and quinonoid intermediates. Mutation of this residue to cysteine would be expected to dramatically impair catalysis, as cysteine cannot occupy the same space as lysine and also exhibits differing pK_a properties. Upon treatment with 2-bromoethylamine, however, the resulting thioether now has a shape and pK_a similar to the original lysine side chain. Hence, some catalytic activity is restored.

26. An external aldimine forms with SAM, which is deprotonated to form the quinonoid intermediate. The deprotonated carbon atom attacks the carbon atom adjacent to the sulfur atom to form the cyclopropane ring and release methylthioadenosine, the other product.

27. An external aldimine forms with L-serine, which is deprotonated to form the quinonoid intermediate. This intermediate is reprotonated on its opposite face to form an aldimine with D-serine. This compound is cleaved to release D-serine. The equilibrium constant for a racemization reaction is 1 because the reactant and product are exact mirror images of each other.

28. (a) In the first step, histidine attacks the methylene group from the methionine subgroup of SAM (rather than the usual methyl substituent), resulting in the transfer of an aminocarboxypropyl group. Three subsequent conventional SAM-mediated methylations of the primary amine yield diphthine.



(b) In this chapter, we have observed two examples of an ATP-dependent conversion of a carboxylate into an amide: glutamine synthetase, which uses an acyl-phosphate intermediate, and asparagine synthetase, which uses an acyl-adenylate intermediate. Either mechanism is possible in formation of diphthamide from diphthine.

29. Biotin.

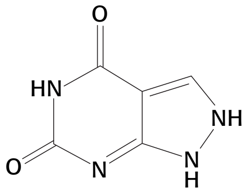
30. Synthesis from oxaloacetate and α-ketoglutarate would deplete the citric acid cycle, which would decrease ATP production. Anapleurotic reactions would be required to replenish the citric acid cycle.

31. SAM is the donor for DNA methylation reactions that protect a host from digestion by its own restriction enzymes. A lack of SAM would render the bacterial DNA susceptible to digestion by the cell's own restriction enzymes.

32. Acetate → acetyl-CoA → citrate → isocitrate → α-ketoglutarate → succinyl-CoA.
33. The value of K_M of glutamate dehydrogenase for NH_4^+ is high (>> 1 mM), and so this enzyme is not saturated when NH_4^+ is limiting. In contrast, glutamine synthetase has very high affinity for NH_4^+. Thus, ATP hydrolysis is required to capture ammonia when it is scarce.
34. (a) Asparagine is much more abundant in the dark. More glutamine is present in the light. These amino acids show the most dramatic effects. Glycine also is more abundant in the light.
(b) Glutamine is a more metabolically reactive amino acid, used in the synthesis of many other compounds. Consequently, when energy is available as light, glutamine will be preferentially synthesized. Asparagine, which carries more nitrogen per carbon atom and is thus a more-efficient means of storing nitrogen when energy is short, is synthesized in the dark. Glycine is more prevalent in the light because of photorespiration.
(c) White asparagus has an especially high concentration of asparagine, which accounts for its intense taste. All asparagus has a large amount of asparagine. In fact, as suggested by its name, asparagine was first isolated from asparagus.

Chapter 25

1. In de novo synthesis, the nucleotides are synthesized from simpler precursor compounds, in essence from scratch. In salvage pathways, preformed bases are recovered and attached to riboses.
2. Carbon 2 and nitrogen 3 come from carbamoyl phosphate. Nitrogen 1 and carbons 4, 5, and 6 are derived from aspartate.
3. Nitrogen 1: aspartate; carbon 2: N^{10}-formyltetrahydrofolate; nitrogen 3: glutamine; carbons 4 and 5 and nitrogen 7: glycine; carbon 6: CO_2; carbon 8: N^{10}-formyltetrahydrofolate; nitrogen 9: glutamine.
4. Energy currency: ATP; signal transduction: ATP and GTP; RNA synthesis: ATP, GTP, CTP, and UTP; DNA synthesis: dATP, dCTP, dGTP, and TTP; components of coenzymes: ATP in CoA, FAD, and $NAD(P)^+$; carbohydrate synthesis: UDP-glucose. These are just some of the uses.
5. A nucleoside is a base attached to ribose. A nucleotide is a nucleoside with the ribose bearing one or more phosphates.
6. (a) 9; (b) 7; (c) 6; (d) 10; (e) 2; (f) 4; (g) 1; (h) 11; (i) 8; (j) 3; (k) 5.
7. Substrate channeling is the process whereby the product of one active site moves to become a substrate at another active site without ever leaving the enzyme. A channel connects the active sites. Substrate channeling greatly enhances enzyme efficiency and minimizes the diffusion of a substrate to an active site.
8. $\text{Glucose} + 2\,\text{ATP} + 2\,\text{NADP}^+ + H_2O \longrightarrow \text{PRPP} + CO_2 + \text{ADP} + \text{AMP} + 2\,\text{NADPH} + 3\,H^+$.
9. $\text{Glutamine} + \text{aspartate} + CO_2 + 2\,\text{ATP} + \text{NAD}^+ \longrightarrow \text{orotate} + 2\,\text{ADP} + 2\,P_i + \text{glutamate} + \text{NADH} + H^+$.
10. (a, c, and d) PRPP; (b) carbamoyl phosphate.
11. PRPP and formylglycinamide ribonucleotide
12. $\text{dUMP} + \text{serine} + \text{NADPH} + H^+ \longrightarrow \text{TMP} + \text{NADP}^+ + \text{glycine}$.
13. There is a deficiency of N^{10}-formyltetrahydrofolate. Sulfanilamide inhibits the synthesis of folate by acting as an analog of *p*-aminobenzoate, one of the precursors of folate.
14. (a) Cell A cannot grow in a HAT medium, because it cannot synthesize TMP either from thymidine or from dUMP. Cell B cannot grow in this medium, because it cannot synthesize purines by either the de novo pathway or the salvage pathway. Cell C can grow in a HAT medium because it contains active thymidine kinase from cell B (enabling it to phosphorylate thymidine to TMP) and hypoxanthine guanine phosphoribosyltransferase from cell A (enabling it to synthesize purines from hypoxanthine by the salvage pathway).
(b) Transform cell A with a plasmid containing foreign genes of interest and a functional thymidine kinase gene. The only cells that will grow in a HAT medium are those that have acquired a thymidylate kinase gene; nearly all of these transformed cells will also contain the other genes on the plasmid.
15. The folate derivative N^5, N^{10}-methylenetetrahydrofolate is required by thymidylate synthase to add a methyl group to dUMP, forming TMP. Insufficient folate could result in spina bifida.
16. The reciprocal substrate relation refers to the fact that AMP synthesis requires GTP, whereas GMP synthesis requires ATP. These requirements tend to balance the synthesis of ATP and GTP.
17. Ring carbon 6 in cytosine will be labeled. In guanine, only carbon 5 will be labeled with ^{13}C.
18. UTP is first converted into UDP. Ribonucleotide reductase generates dUDP. DeoxyUDP is converted to dUMP. Thymidylate synthase generates TMP from dUMP. Monophosphate and diphosphate kinases subsequently form TTP.
19. These patients have a high level of urate because of the breakdown of nucleic acids. Allopurinol prevents the formation of kidney stones and blocks other deleterious consequences of hyperuricemia by preventing the formation of urate.
20. The free energies of binding are -57.7 (wild type), -49.8 (Asn 27), and -38.1 (Ser 27) kJ mol^{-1} (-13.8, -11.9, and -9.1 kcal mol^{-1}, respectively). The loss in binding energy is $+7.9$ kJ mol^{-1} ($+1.9$ kcal mol^{-1}) and $+19.7$ kJ mol^{-1} ($+4.7$ kcal mol^{-1}).
21. By their nature, cancer cells divide rapidly and thus require frequent DNA synthesis. Inhibitors of TMP synthesis will impair DNA synthesis and cancer growth.
22. Uridine and cytidine are administered to by-pass the defective enzyme in orotic aciduria.
23. Inosine or hypoxanthine could be administered.
24. N-1 in both cases, and the amine group linked to C-6 in ATP.
25. Nitrogen atoms 3 and 9 in the purine ring
26. Allopurinol, an analog of hypoxanthine, is a suicide inhibitor of xanthine oxidase.
27. An oxygen atom is added to allopurinol to form alloxanthine.



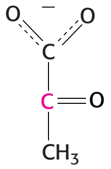
Alloxanthine

28.

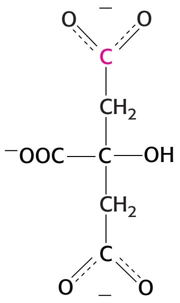
The synthesis of carbamoyl phosphate requires 2 ATP	2 ATP
The formation of PRPP from ribose 5-phosphate yields an AMP*	2 ATP
The conversion of UMP to UTP requires 2 ATP	2 ATP
The conversion of UTP to CTP requires 1 ATP	1 ATP
Total	7 ATP

*Remember that AMP is the equivalent of 2 ATP because an ATP must be expended to generate ADP, the substrate for ATP synthesis.

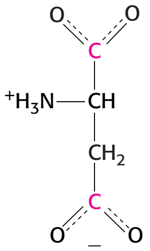
29. (a) Carboxyaminoimidazole ribonucleotide; (b) glycinamide ribonucleotide; (c) phosphoribosyl amine; (d) formylglycinamide ribonucleotide.
30. The first reaction proceeds by phosphorylation of glycine to form an acyl phosphate followed by nucleophilic attack by the amine of phosphoribosylamine to displace orthophosphate. The second reaction consists of adenylation of the carbonyl group of xanthylate followed by nucleophilic attack by ammonia to displace AMP.
31. The $-NH_2$ group attacks the carbonyl carbon atom to form a tetrahedral intermediate. Removal of a proton leads to the elimination of water to form inosinate.
32. The enzyme that uses ammonia, carbamoyl phosphate synthetase I, forms carbamoyl phosphate for a reaction with ornithine, the first step of the urea cycle. The enzyme that uses glutamine, carbamoyl phosphate synthetase II, generates carbamoyl phosphate for use in the first step of pyrimidine biosynthesis.
33. PRPP is the activated intermediate in the synthesis of phosphoribosylamine in the de novo pathway of purine formation; of purine nucleotides from free bases by the salvage pathway; of orotidylate in the formation of pyrimidines; of nicotinate ribonucleotide; of phosphoribosyl ATP in the pathway leading to histidine; and of phosphoribosylanthranilate in the pathway leading to tryptophan.
34. (a) cAMP; (b) ATP; (c) UDP-glucose; (d) acetyl CoA; (e) NAD^+, FAD; (f) dideoxynucleotides; (g) fluorouracil; (h) CTP inhibits ATCase.
35. In vitamin B_{12} deficiency, methyltetrahydrofolate cannot donate its methyl group to homocysteine to regenerate methionine. Because the synthesis of methyltetrahydrofolate is irreversible, the cell's tetrahydrofolate will ultimately be converted into this form. No formyl or methylene tetrahydrofolate will be left for nucleotide synthesis. Vitamin B_{12} is also required to metabolize propionyl CoA generated in the oxidation of odd-chain fatty acids and in the degradation of methionine.
36. Because folate is required for nucleotide synthesis, cells that are dividing rapidly would be most readily affected. They would include cells of the intestine, which are constantly replaced, and precursors to blood cells. A lack of intestinal cells and blood cells would account for the symptoms often observed.
37. In patients with glucose 6-phosphatase deficiency, the cytoplasmic level of ATP in the liver falls as a result of increased glycogenolysis. In all three conditions, AMP rises above normal, and the excess AMP is degraded to urate.
38. Succinate → malate → oxaloacetate by the citric acid cycle. Oxaloacetate → aspartate by transamination, followed by pyrimidine synthesis. Carbons 4, 5, and 6 are labeled.
39. Glucose will most likely be converted into two molecules of pyruvate, one of which will be labeled in the 2 position:



Now consider two common fates of pyruvate—conversion into acetyl CoA and subsequent processing by the citric acid cycle or carboxylation by pyruvate carboxylase to form oxaloacetate. Formation of citrate by condensing the labeled pyruvate with oxaloacetate will yield labeled citrate:



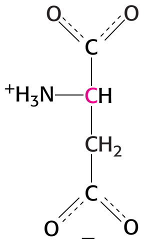
The labeled carbon will be retained through one round of the citric acid cycle but, on the formation of the symmetric succinate, the label will appear in two different positions. Thus, when succinate is metabolized to oxaloacetate, which may be aminated to form aspartate, two carbons will be labeled:



When this aspartate is used to form uracil, the labeled COO^- attached to the α-carbon is lost and the other COO^- becomes incorporated into uracil as carbon 4.

Suppose, instead, that labeled 2-[^{14}C]pyruvate is carboxylated to form oxaloacetate and processed to form aspartate. In this case, the α-carbon of aspartate bears the label.

When this aspartate is used to synthesize uracil, carbon 6 bears the label:



40. HGPRT, which is nonfunctional in Lesch–Nyhan patients, is required to form 6-mercaptopurine ribose monophosphate. Consequently, de novo purine synthesis continues.
41. (a) Some ATP can be salvaged from the ADP that is being generated. (b) There are equal numbers of high-phosphoryl-transfer-potential groups on each side of the equation. (c) Because the adenylate kinase reaction is at equilibrium, the removal of AMP would lead to the formation of more ATP. (d) Essentially, the cycle serves as an anaplerotic reaction for the generation of the citric acid cycle intermediate fumarate.
42. (i) The formation of 5-aminoididazole-4-carboxamide ribonucleotide from 5-aminoimidazole-4-(*N*-succinylcarboxamide) ribonucleotide in the synthesis of IMP. (ii) The formation of AMP from adenylosuccinate. (iii) The formation of arginine from argininosuccinate in the urea cycle.
43. Allopurinol is an inhibitor of xanthine oxidase, which is on the pathway for urate synthesis. In your pet duck, this pathway is the

means by which excess nitrogen is excreted. If xanthine oxidase were inhibited in your duck, nitrogen could not be excreted, with severe consequences such as the formation of a dead duck.

44. (a) The enzyme from the LND 2 patient showed much less activity than the normal enzyme, suggesting that the defect in the enzyme impaired its catalytic ability. The results for LND 1 are puzzling. The enzyme displays activity similar to the enzyme from the normal cell line, and yet the patient was suffering from Lesch–Nyhan disease.

(b) Possible explanations include: there may be an inhibitor in the cells that prevents the enzyme from acting in vivo but is lost during the purification procedure; the enzyme may be degraded more rapidly in vivo than the normal enzyme; the enzyme may be inherently less stable than the normal enzyme.

(c) The enzyme from LND 1 lost activity much faster than the normal enzyme, suggesting that the enzyme was structurally unstable and would lose enzyme activity in the cell, accounting for the appearance of the disease.

Chapter 26

1. Glycerol 3-phosphate is the foundation for both triacylglycerol and phospholipid synthesis. Glycerol 3-phosphate is acylated twice to form phosphatidate. In triacylglycerol synthesis, the phosphoryl group is removed from glycerol 3-phosphate to form diacylglycerol, which is then acylated to form triacylglycerol. In phospholipid synthesis, phosphatidate commonly reacts with CTP to form CDP-diacylglycerol, which then reacts with an alcohol to form a phospholipid. Alternatively, diacylglycerol may react with a CDP-alcohol to form a phospholipid.

2. Glycerol 3-phosphate is formed primarily by the reduction of dihydroxyacetone phosphate, a gluconeogenic intermediate, and to a lesser extent by the phosphorylation of glycerol.

3. Glycerol + 4 ATP + 3 fatty acids + 4 $H_2O \longrightarrow$ triacylglycerol + ADP + 3 AMP + 7 P_i + 4 H^+.

4. Glycerol + 3 ATP + 2 fatty acids + 2 H_2O + CTP + ethanolamine $\longrightarrow$ phosphatidylethanolamine + CMP + ADP + 2 AMP + 6 P_i + 3 H^+.

5. Three. One molecule of ATP to form phosphorylethanolamine and two molecules of ATP to regenerate CTP from CMP.

6. All are synthesized from ceramide. In sphingomyelin, the terminal hydroxyl group of ceramide is modified with phosphorylcholine. In a cerebroside, the hydroxyl group has a glucose or galactose attached. In a ganglioside, oligosaccharide chains are attached to the hydroxyl group.

7. (i) Activate the diacylglycerol as CDP-DAG. (ii) Activate the alcohol as CDP-alcohol. (iii) Use the base-exchange reaction.

8. (a) CDP-diacylglycerol; (b) CDP-ethanolamine; (c) acyl CoA; (d) phosphatidylcholine; (e) UDP-glucose or UDP-galactose; (f) UDP-galactose; (g) geranyl pyrophosphate.

9. Such mutations are seen in mice. The amount of adipose tissue would decrease severely because diacylglycerol could not be formed. Normally, diacylglycerol is acylated to form triacylglycerols. If there were deficient phosphatidic acid phosphatase activity, no triacylglycerols would form.

10. (a) 8; (b) 4; (c) 1; (d) 9; (e) 3; (f) 10; (g) 5; (h) 2; (i) 6; (j) 7.

11. (i) The synthesis of activated isoprene units (isopentyl pyrophosphate), (ii) the condensation of six of the activated isoprene units to form squalene, and (iii) cyclization of the squalene to form cholesterol.

12. The amount of reductase and its activity control the regulation of cholesterol biosynthesis. Transcriptional control is mediated by SREBP. Translation of the reductase mRNA also is controlled. The reductase itself may undergo regulated proteolysis. Finally, the activity of the reductase is inhibited by phosphorylation by AMP kinase when ATP levels are low.

13. (a and b) None, because the label is lost as CO_2.

14. The hallmark of this genetic disease is elevated cholesterol levels in the blood of even young children. The excess cholesterol is taken up by marcrophages, which eventually results in the formation of plaques and heart disease. There are many mutations that cause the disease, but all result in malfunctioning of the LDL receptor.

15. The categories of mutations are: (i) no receptor is synthesized; (ii) receptors are synthesized but do not reach the plasma membrane, because they lack signals for intracellular transport or do not fold properly; (iii) receptors reach the cell surface, but they fail to bind LDL normally because of a defect in the LDL-binding domain; (iv) receptors reach the cell surface and bind LDL, but they fail to cluster in coated pits because of a defect in their carboxyl-terminal regions.

16. "None of your business" and "I don't talk biochemistry until after breakfast" are appropriate but rude and uninformative answers. A better answer might be: "Although it is true that cholesterol is a precursor to steroid hormones, the rest of the statement is oversimplified. Cholesterol is a component of membranes, and membranes literally define cells, and cells make up tissues. But to say that cholesterol 'makes' cells and tissues is wrong."

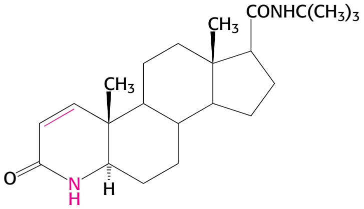
17. Statins are competitive inhibitors of HMG-CoA reductase. They are used as drugs to inhibit cholesterol synthesis in patients with high levels of cholesterol.

18. No. Cholesterol is essential for membrane function and as a precursor for bile salts and steroid hormones. The complete lack of cholesterol would be lethal.

19. Deamination of cytidine to uridine changes CAA (Gln) into UAA (stop).

20. The LDL contains apolipoprotein B-100, which binds to an LDL receptor on the cell surface in a region known as a coated pit. On binding, the complex is internalized by endocytosis to form an internal vesicle. The vesicle is separated into two components. One, with the receptor, is transported back to the cell surface and fuses with the membrane, allowing continued use of the receptor. The other vesicle fuses with lysosomes inside the cell. The cholesteryl esters are hydrolyzed, and free cholesterol is made available for cellular use. The LDL protein is hydrolyzed to free amino acids.

21. Benign prostatic hypertrophy can be treated by inhibiting 5α-reductase. Finasteride, the 4-azasteroid analog of dihydrotestosterone, competitively inhibits the reductase but does not act on androgen receptors. Patients taking finasteride have a markedly lower plasma level of dihydrotestosterone and a nearly normal level of testosterone. The prostate gland becomes smaller, whereas testosterone-dependent processes such as fertility, libido, and muscle strength appear to be unaffected.



Finasteride

22. Patients who are most sensitive to debrisoquine have a deficiency of a liver P450 enzyme encoded by a member of the *CYP2* subfamily. This characteristic is inherited as an autosomal recessive

trait. The capacity to degrade other drugs may be impaired in people who hydroxylate debrisoquine at a slow rate, because a single P450 enzyme usually handles a broad range of substrates.

23. Many hydrophobic odorants are deactivated by hydroxylation. Molecular oxygen is activated by a cytochrome P450 monooxygenase. NADPH serves as the reductant. One oxygen atom of O_2 goes into the odorant substrate, whereas the other is reduced to water.

24. Recall that dihydrotestosterone is crucial for the development of male characteristics in the embryo. If a pregnant woman were to be exposed to Propecia, the 5α-reductase of the male embryo would be inhibited, which could result in severe developmental abnormalities.

25. The oxygenation reactions catalyzed by the cytochrome P450 family permit greater flexibility in biosynthesis. Because plants are not mobile, they must rely on physical defenses, such as thorns, and chemical defenses, such as toxic alkaloids. The larger P450 array might permit greater biosynthetic versatility.

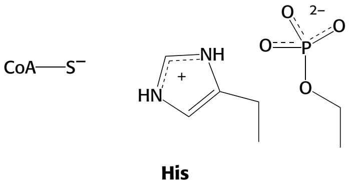
26. This knowledge might enable clinicians to characterize the likelihood of a patient's having an adverse drug reaction or being susceptible to chemical-induced illnesses. It might also permit a personalized and especially effective drug-treatment regime for diseases such as cancer.

27. The honeybees may be especially sensitive to environmental toxins, including pesticides, because these chemicals are not readily detoxified, owing to the minimal P450 system.

28. A deficiency of the hydroxylase will cause a build-up of progesterone, which will then be converted into estradiol and testosterone.

29. The core structure of a steroid is four fused rings: three cyclohexane rings and one cyclopentane ring. In vitamin D, the B ring is split by ultraviolet light.

30. The negatively charged phosphoserine residue interacts with the positively charged protonated histidine residue and decreases its ability to transfer a proton to the thiolate.



31. The methyl group is first hydroxylated. The hydroxymethylamine eliminates formaldehyde to form methylamine.

32. Note that a cytidine nucleotide plays the same role in the synthesis of these phosphoglycerides as a uridine nucleotide does in the formation of glycogen (Section 21.4). In all of these biosyntheses, an activated intermediate (UDP-glucose, CDP-diacylglycerol, or CDP-alcohol) is formed from a phosphorylated substrate (glucose 1-phosphate, phosphatidate, or a phosphorylalcohol) and a nucleoside triphosphate (UTP or CTP). The activated intermediate then reacts with a hydroxyl group (the terminus of glycogen, the side chain of serine, or a diacylglycerol).

33. The attachment of isoprenoid side chains confers hydrophobic character. Proteins having such a modification are targeted to membranes.

34. 3-Hydroxy-3-methylglutaryl CoA is also a precursor for ketone-body synthesis. If fuel is needed elsewhere in the body, as might be the case during a fast, 3-hydroxy-3-methylglutaryl CoA is converted into the ketone body acetoacetate. If energy needs are met, the liver will synthesize cholesterol. Remember also that cholesterol synthesis occurs in the cytoplasm while ketone bodies are synthesized in the mitochondria.

35. One way in which phosphatidylcholine can be synthesized is by the addition of three methyl groups to phosphatidylethanolamine. The methyl donor is a modified form of methionine, *S*-adenosylmethionine or SAM (Section 24.2).

36. Mutations could occur in the gene encoding the sodium channel that would prevent the action of DDT. Alternatively, P450 enzyme synthesis could be increased to accelerate metabolism of the insecticide to inactive metabolites. In fact, both types of responses to insecticides have been observed.

37. Citrate is transported out of the mitochondria in times of plenty. ATP-citrate lyase yields acetyl CoA and oxaloacetate. The acetyl CoA can then be used to synthesize cholesterol.

38. (a) There is no effect. (b) Because actin is not controlled by cholesterol, the amount isolated should be the same in both experimental groups; a difference would suggest a problem in the RNA isolation. (c) The presence of cholesterol in the diet dramatically reduces the amount of HMG-CoA reductase protein. (d) A common means of regulating the amount of a protein present is to regulate transcription, which is clearly not the case here. (e) The translation of mRNA could be inhibited, and the protein could be rapidly degraded.

39. (a) The activity of the enzyme from the GD patient is the same as that from the control. This suggests that GD is not caused by a mutation that impairs enzyme activity.

(b) There is much less GCase in the GD sample. Because the same number of cells were used in the two samples (confirmed by the identical levels of actin in both samples), we know that the difference in amount of GCase is not due to different quantities of cells. Two possible explanations are that transcription is impaired in the GD sample or that the enzyme in GD is more readily destroyed than the control.

(c) Apparently, the enzyme is being synthesized but then degraded.

(d) The defect in the enzyme from the GD patient seems to result in more rapid degradation of the enzyme by the proteasome. The fact that the amount of GCase present in the control increased upon proteasomal inhibition suggests that GCase may undergo rapid turnover under normal conditions.

Chapter 27

1. Over the 40 years under consideration, our test subject will have consumed

$$\begin{aligned}40 \text{ years} \times 365 \text{ days year}^{-1} \times 8400 \text{ kJ } (2000 \text{ kcal}) \text{ day}^{-1} \\ = 1.2 \times 10^8 \text{ kJ } (2.9 \times 10^7 \text{ kcals}) \text{ in } 40 \text{ years}\end{aligned}$$

Thus, over the 40-year span, our subject has ingested

$$\begin{aligned}1.2 \times 108(2.9 \times 10^7 \text{ kcal})/38 \text{ kJ } (9 \text{ kcal g}^{-1}) \\ = 3.2 \times 10^6 \text{ g} = 3200 \text{ kg of food}\end{aligned}$$

which is equivalent to more than 6 tons of food!

2. 55 pounds = 25 kg = 25,000 g = total weight gain

$40 \text{ years} \times 365 \text{ days year}^{-1} = 14{,}600 \text{ days}$

$25{,}000 \text{ g}/14{,}600 \text{ days} = 1.7 \text{ g day}^{-1}$

which is equivalent to an extra pat of butter per day. Her BMI is 26.5, and she would be considered overweight but not obese.

3. Adipose tissue is now known to be an active endocrine organ, secreting signal molecules called adipokines.

4. Caloric homeostasis is the condition in which the energy expenditure of an organism is equal to the energy intake.

5. Leptin and insulin.

6. CCK produces a feeling of satiety and stimulates the secretion of digestive enzymes by the pancreas and the secretion of bile salts

by the gall bladder. GLP-1 also produces a feeling of satiety; in addition, it potentiates the glucose-induced secretion of insulin by the β cells of the pancreas.

7. Obviously, something is amiss. Although the answer is not known, the leptin-signaling pathway appears to be inhibited by suppressors of cytokine signaling, the regulatory proteins.

8. 1: a, b; 2: f; 3: c, d, f; 4: c, d; 5: c; 6: f; 7: e; 8: e; 9: e.

9. Phosphorylation of dietary glucose after it enters the liver; gluconeogenesis; glycogen breakdown.

10. Type 1 diabetes is due to autoimmune destruction of the insulin-producing cells of the pancreas. Type 1 diabetes is also called insulin-dependent diabetes because affected people require insulin to survive. Type 2 diabetes is characterized by insulin resistance. Insulin is produced, but the tissues that should respond to insulin, such as muscle, do not.

11. Leptin stimulates processes impaired in diabetes. For instance, leptin stimulates fatty acid oxidation, inhibits triacylglycerol synthesis, and increases the sensitivity of muscle and the liver to insulin.

12. (a) A watt is equal to 1 joule (J) per second (0.239 calorie per second). Hence, 70 W is equivalent to 0.07 kJ s^{-1} (0.017 kcal s^{-1}).
(b) A watt is a current of 1 ampere (A) across a potential of 1 volt (V). For simplicity, let us assume that all the electron flow is from NADH to O_2 (a potential drop of 1.14 V). Hence, the current is 61.4 A, which corresponds to 3.86×10^{20} electrons per second (1 A = 1 coulomb s^{-1} = 6.28×10^{18} charge s^{-1}).
(c) About 2.5 molecules of ATP are formed per molecule of NADH oxidized (two electrons). Hence, 1 molecule of ATP is formed per 0.8 electron transferred. A flow of 3.86×10^{20} electrons per second therefore leads to the generation of 4.83×10^{20} molecules of ATP per second, or 0.80 mmol s^{-1}.
(d) The molecular weight of ATP is 507. The total body content of ATP of 50 g is equal to 0.099 mol. Hence, ATP turns over about once in 125 seconds when the body is at rest.

13. (a) The stoichiometry of the complete oxidation of glucose is

$$C_6H_{12}O_6 + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O$$

and that of tripalmitoylglycerol is

$$C_{51}H_{98}O_2 + 72.5\,O_2 \longrightarrow 51\,CO_2 + 49\,H_2O$$

Hence, the RQ values are 1.0 and 0.703, respectively.
(b) An RQ value reveals the relative use of carbohydrates and fats as fuels. The RQ of a marathon runner typically decreases from 0.97 to 0.77 in the course of a race. The lowering of the RQ indicates the shift in fuel from carbohydrates to fat.

14. One gram of glucose (molecular weight 180.2) is equal to 5.55 mmol, and one gram of tripalmitoylglycerol (molecular weight 807.3) is equal to 1.24 mmol. The reaction stoichiometries (Problem 13) indicate that 6 mol of H_2O is produced per mole of glucose oxidized, and 49 mol of H_2O is produced per mole of tripalmitoylglycerol oxidized. Hence, the H_2O yields per gram of fuel are 33.3 mmol (0.6 g) for glucose and 60.8 mmol (1.09 g) for tripalmitoylglycerol. Thus, complete oxidation of this fat gives 1.82 times as much water as does glucose. Another advantage of triacylglycerols is that they can be stored in essentially anhydrous form, whereas glucose is stored as glycogen, a highly hydrated polymer. A hump consisting mainly of glycogen would be an intolerable burden—far more than the straw that broke the camel's back.

15. The starved–fed cycle is the nightly hormonal cycle that humans experience during sleep and on eating. The cycle maintains adequate amounts of blood glucose. The starved part—sleep—is characterized by increased glucagon secretion and decreased insulin secretion. After a meal, glucagon concentration falls and insulin concentration rises.

16. Ethanol is oxidized to yield acetaldehyde by alcohol dehydrogenase, which is subsequently oxidized to acetate and acetaldehyde. Ethanol is also metabolized to acetaldehyde by the P450 enzymes, with the subsequent depletion of NADPH.

17. First, fatty liver develops owing to the increased amounts of NADH that inhibit fatty acid oxidation and stimulate fatty acid synthesis. Second, alcoholic hepatitis begins owing to oxidative damage and damage due to excess acetaldehyde that results in cell death. Finally, fibrous tissues form, creating scars that impair blood flow and biochemical function. Ammonia cannot be converted into urea, and its toxicity leads to coma and death.

18. A typical macadamia nut has a mass of about 2 g. Because it consists mainly of fats (~37 kJ g^{-1}, ~9 kcal g^{-1}), a nut has a value of about 75 kJ (18 kcal). The ingestion of 10 nuts results in an intake of about 753 kJ (180 kcal). As stated in the answer to Problem 12, a power consumption of 1 W corresponds to 1 J s^{-1} (0.239 cal s^{-1}), and so 400-W running requires 0.4 kJ s^{-1} (0.0956 kcal s^{-1}). Hence, a person would have to run 1882 s, or about 31 minutes, to spend the calories provided by 10 nuts.

19. A high blood-glucose level triggers the secretion of insulin, which stimulates the synthesis of glycogen and triacylglycerols. A high insulin level would impede the mobilization of fuel reserves during the marathon.

20. A lack of adipose tissue leads to an accumulation of fats in the muscle, with the generation of insulin resistance. The experiment shows that adipokines secreted by the adipose tissue, here leptin, facilitate in some fashion the action of insulin in muscle.

21. Such a mutation would increase the phosphorylation of the insulin receptor and IRS in muscle and would improve insulin sensitivity. Indeed, PTP1B is an attractive therapeutic target for type 2 diabetes.

22. Lipid mobilization can be so rapid that it exceeds the ability of the liver to oxidize the lipids or convert them into ketone bodies. The excess is reesterified and released into the blood as VLDLs.

23. A role of the liver is to provide glucose for other tissues. In the liver, glycolysis is used not for energy production but for biosynthetic purposes. Consequently, in the presence of glucagon, liver glycolysis stops so that the glucose can be released into the blood.

24. The urea cycle and gluconeogenesis.

25. (a) Insulin inhibits lipid utilization.
(b) Insulin stimulates protein synthesis, but there are no amino acids in the children's diet. Moreover, insulin inhibits protein breakdown. Consequently, muscle proteins cannot be broken down and used for the synthesis of essential proteins.
(c) Because proteins cannot be synthesized, blood osmolarity is too low. Consequently, fluid leaves the blood. An especially important protein for maintaining blood osmolarity is albumin.

26. During strenuous exercise, muscle converts glucose into pyruvate through glycolysis. Some of the pyruvate is processed by cellular respiration. However, some of it is converted into lactate and released into the blood. The liver takes up the lactate and converts it into glucose through gluconeogenesis. Muscle may process the carbon skeletons of branched-chain amino acids aerobically. The nitrogens of these amino acids are transferred to pyruvate to form alanine, which is released into the blood and taken up by the liver. After the transamination of the amino group to α-ketoglutarate, the resulting pyruvate is converted into glucose. Finally, muscle glycogen may be mobilized, and the released glucose can be used by muscle.

27. This conversion allows muscle to function anaerobically. NAD^+ is regenerated when pyruvate is reduced to lactate, and so energy can continue to be extracted from glucose during strenuous exercise. The liver converts the lactate into glucose.

28. Fatty acids and glucose, respectively.

29. This practice is called carbo-loading. Depleting the glycogen stores will initially cause the muscles to synthesis a large amount of glycogen when dietary carbohydrates are provided and will lead to the supercompensation of glycogen stores.

30. The oxygen consumption at the end of exercise is used to replenish ATP and creatine phosphate and to oxidize any lactate produced.

31. Oxygen is used in oxidative phosphorylation to resynthesize ATP and creatine phosphate. The liver converts lactate released by the muscle into glucose. Blood must be circulated to return the body temperature to normal, and so the heart cannot return to its resting rate immediately. Hemoglobin must be reoxygenated to replace the oxygen used in exercise. The muscles that power breathing must continue working at the same time as the exercised muscles are returning to resting states. In essence, all the biochemical systems activated in intense exercise need increased oxygen to return to the resting state.

32. Ethanol may replace water that is hydrogen bonded to proteins and membrane surfaces. This alteration of the hydration state of the protein would alter its conformation and hence function. Ethanol may also alter phospholipid packing in membranes. The two effects suggest that integral membrane proteins would be most sensitive to ethanol, as indeed seems to be the case.

33. Cells from the type I fiber would be rich in mitochondria, whereas those of the type II fiber would have few mitochondria.

34. (a) The ATP expended during this race amounts to about 8380 kg, or 18,400 pounds. (b) The cyclist would need about $1,260,000,000 to complete the race.

35. Exercise greatly enhances the ATP needs of muscle cells. To more efficiently meet these needs, more mitochondria are synthesized.

36. AMPK activity increases as ATP is used for muscle contraction. AMPK inactivates acetyl CoA carboxylase. Recall that malonyl CoA, the product of acetyl CoA carboxylase, inhibits transport of fatty acids into the mitochondria. The decrease in muscle malonyl CoA allows fatty acid oxidation in muscle.

37. After the first several days of starvation, most tissues are using fatty acids. Ketone bodies provide much of the brain's energy, decreasing glucose requirement. The low insulin/high glucagon ration stimulates lipolysis and gluconeogenesis. However, gluconeogenesis will not occur because the increase in fatty acid oxidation increases acetyl CoA and NADH. The high concentration of NADH inhibits the TCA cycle as well as gluconeogenesis. The acetyl CoA forms ketone bodies.

38. (a) The increase in all of the ratios is due to the NADH glut caused by the metabolism of alcohol.
(b) The increased amounts of lactate and D-3-hydroxybutyrate are released into the blood, accounting for the acidosis.
(c) Drinking on an empty stomach suggests that glycogen stores are low. Because of the NADH glut, gluconeogenesis cannot occur. Consequently, hypoglycemia results. A well-fed drinker will have glucose in the blood from the meal and thus will not experience hypoglycemia.

39. If glucose is always provided, even in small amounts, the brain will use glucose as a fuel rather than adapting to ketone body use. During the fast, muscle protein will be broken down to meet the brains glucose needs. This protein degradation will lead to organ failure sooner than if the brain had adapted to ketone body utilization.

40. The inability of muscle mitochondria to process all of the fatty acids produced by overnutrition leads to excessive levels of diacylglycerol and ceramide in the muscle cytoplasm. These second-messenger molecules activate enzymes that impair insulin signaling.

41. Both are due to a lack of thiamine (vitamin B_1). Thiamine, which is sometimes called aneurin, is required most notably for the proper functioning of pyruvate dehydrogenase.

42. (a) Red blood cells always produce lactate, and fast-twitch or type II, muscle fibers (Problem 33) also produce a large amount of lactate.
(b) At that point, the athlete is beginning to move into anaerobic exercise, in which most energy is produced by anaerobic glycolysis.
(c) The lactate threshold is essentially the point at which the athlete switches from aerobic exercise, which can be done for extended periods, to anaerobic exercise, essentially sprinting, which can be done for only short periods. The idea is to race at the extreme of his or her aerobic capacity until the finish line is in sight and then to switch to anaerobic.
(d) Training increases the amount of blood vessels and the number of muscle mitochondria. Together, they increase the ability to process glucose aerobically. Consequently, a greater effort can be expended before the switch to anaerobic energy production.

43. Consider the graph of lactate production as a function of effort shown in Problem 42. For an athlete racing at her lactate threshold or just below it, the RQ value will be 1. With the finish line in sight, our runner ups her pace so that she is now also processing glucose to lactic acid in addition to processing glucose aerobically; that is, she begins running above her lactate threshold. The lactic acid released into the blood will ionize

$$\underset{\text{Lactic acid}}{\mathrm{CH_3CH(OH)COOH}} \rightleftharpoons \mathrm{CH_3CH(OH)COO^-}$$

The increase in H^+ will alter the blood buffer system, leading to the formation of carbonic acid

$$\mathrm{H^+ + HCO_3^-} \rightleftharpoons \underset{\text{Carbonic acid}}{\mathrm{H_2CO_3}}$$

The carbonic acid will dissociate into water and carbon dioxide

$$\mathrm{H_2CO_3} \rightleftharpoons \mathrm{CO_2 + H_2O}$$

This carbon dioxide will be superimposed on the carbon dioxide generated by the combustion of glucose aerobically, leading to an RQ greated than 1.

44. The increase in ATP resulting from the processing of glucose in the β cells closes a potassium channel. The closing of the potassium channel alters the voltage across the cell membrane, which leads to an opening of a calcium channel. The influx of calcium causes the insulin-containing secretory granules to fuse with the cell membrane, resulting in the release of insulin.

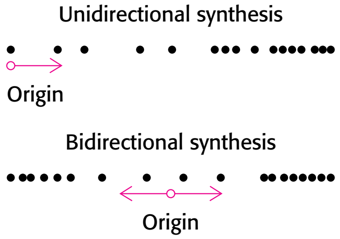
Chapter 28

1. DNA polymerase I uses deoxyribonucleoside triphosphates; pyrophosphate is the leaving group. DNA ligase uses DNA–adenylate (AMP joined to the 5′-phosphoryl group) as a reaction partner; AMP is the leaving group. Topoisomerase I uses a DNA–tyrosyl intermediate (5′-phosphoryl group linked to the phenolic OH group); the tyrosine residue of the enzyme is the leaving group.

2. Positive supercoiling resists the unwinding of DNA. The melting temperature of DNA increases in proceeding from negatively supercoiled to relaxed to positively supercoiled DNA. Positive supercoiling is probably an adaptation to high temperature.

3. The nucleotides used for DNA synthesis have the triphosphate attached to the 5′-hydroxyl group with free 3′-hydroxyl groups. Such nucleotides can be utilized only for 5′-to-3′ DNA synthesis.

4. DNA replication requires RNA primers. Without appropriate ribonucleotides, such primers cannot be synthesized.
5. This close contact prevents the incorporation of ribonucleotides rather than 2′-deoxyribonucleotides.
6. (a) 96.2 revolutions per second (1000 nucleotides per second divided by 10.4 nucleotides per turn for B-DNA gives 96.2 rps). (b) $0.34\ \mu\text{m s}^{-1}$ (1000 nucleotides per second corresponds to $3400\ \text{Å s}^{-1}$ because the axial distance between nucleotides in B-DNA is 3.4 Å).
7. Eventually, the DNA would become so tightly wound that movement of the replication complex would be energetically impossible.
8. Linking number $Lk = Tw + Wr = 48 + 3 = 51$. If $Tw = 50$, then $Wr = 1$.
9. A hallmark of most cancer cells is prolific cell division, which requires DNA replication. If the telomerase were not activated, the chromosomes would shorten until they became nonfunctional, leading to cell death.
10. No.
11. Treat the DNA briefly with endonuclease to occasionally nick each strand. Add the polymerase with the radioactive dNTPs. At the broken bond, or nick, the polymerase will degrade the existing strand with its $5' \to 3'$ exonuclease activity and replace it with a radioactive complementary copy by using its polymerase activity. This reaction scheme is referred to as "nick translation" because the nick is moved, or translated, along the DNA molecule without ever becoming sealed.
12. If replication were unidirectional, tracks with a low grain density at one end and a high grain density at the other end would be seen. On the other hand, if replication were bidirectional, the middle of a track would have a low density, as shown in the diagram below. For *E. coli,* the grain tracks are denser on both ends than in the middle, indicating that replication is bidirectional.



13. Pro (CCC), Ser (UCC), Leu (CUC), and Phe (UUC). Alternatively, the last base of each of these codons could be U.
14. Potentially deleterious side reactions are prevented. The enzyme itself might be damaged by light if it could be activated by light in the absence of bound DNA harboring a pyrimidine dimer.
15. The free DNA ends that appear in the absence of telomeres are repaired by DNA fusion.
16. The free energy of ATP hydrolysis under standard conditions is $-30.5\ \text{kJ mol}^{-1}$ ($-7.3\ \text{kcal mol}^{-1}$). In principle, it could be used to break three base pairs.
17. The oxidation of guanine could lead to DNA repair: DNA strand cleavage could allow looping out of the triplet repeat regions and triplet expansion.
18. The release of DNA topoisomerase II after the enzyme has acted on its DNA substrate requires ATP hydrolysis. Negative supercoiling requires only the binding of ATP, not its hydrolysis.
19. (a) Size; the top is relaxed and the bottom is supercoiled DNA. (b) Topoisomers. (c) The DNA is becoming progressively more unwound, or relaxed, and thus slower moving.
20. (a) It was used to determine the number of spontaneous revertants—that is, the background mutation rate.
(b) To firmly establish that the system was working. A known mutagen's failure to produce revertants would indicate that something was wrong with the experimental system.
(c) The chemical itself has little mutagenic ability but is apparently activated into a mutagen by the liver homogenate.
(d) Cytochrome P450 system.

Chapter 29

1. The sequence of the coding (+, sense) strand is

5′-ATGGGGAACAGCAAGAGTGGGGCCCTGTCCAAGGAG-3′

and the sequence of the template (–, antisense) strand is

3′-TACCCCTTGTCGTTCTCACCCCGGGACAGGTTCCTC-5′

2. An error will affect only one molecule of mRNA of many synthesized from a gene. In addition, the errors do not become a permanent part of the genomic information.
3. At any given instant, only a fraction of the genome (total DNA) is being transcribed. Consequently, speed is not essential.
4. The active sites are related by convergent evolution.
5. Heparin, a glycosaminoglycan, is highly anionic. Its negative charges, like the phosphodiester bridges of DNA templates, allow it to bind to lysine and arginine residues of RNA polymerase.
6. This mutant σ will competitively inhibit the binding of holoenzyme and prevent the specific initiation of RNA chains at promoter sites.
7. The core enzyme without σ binds more tightly to the DNA template than does the holoenzyme. The retention of σ after chain initiation would make the mutant RNA polymerase slower. The mutant enzyme would also be unlikely to bind alternative σ factors.
8. A 100-kDa protein contains about 910 residues, which are encoded by 2730 nucleotides. At a maximal transcription rate of 50 nucleotides per second, the mRNA would be synthesized in 55 s.
9. The RNA polymerase slides along the DNA rapidly rather than simply diffusing through three-dimensional space.
10. The start site is in red:

5′-GCCGTTGACACCGTTCGGCGATCGATCCGCTATAATGTGTGGATCCGCTT-3′

11. Initiation at strong promoters takes place every 2 s. In this interval, 100 nucleotides are transcribed. Hence, centers of transcription bubbles are 34 nm (340 Å) apart.
12. (a) The lowest band on the gel will be that of strand 3 alone (*i*). Band *ii* will be at the same position as band *i* because the RNA is not complementary to the nontemplate strand, whereas band *iii* will be higher because a complex is formed between RNA and the template strand. Band *iv* will be higher than the others because strand 1 is complexed to 2, and strand 2 is complexed to 3. Band *v* is the highest because core polymerase associates with the three strands.
(b) None, because rifampicin acts before the formation of the open complex.
(c) RNA polymerase is processive. When the template is bound, heparin cannot enter the DNA-binding site.
(d) When GTP is absent, synthesis stops when the first cytosine residue downstream of the bubble is encountered in the template strand. In contrast, with all four nucleoside triphosphates present, synthesis will continue to the end of the template.
13. Backtracking must occur prior to cleavage leading to dinucleotide products.
14. The base-pairing energy of the di- and trinucleotide DNA–RNA hybrids formed at the very beginning of transcription is not sufficient to prevent strand separation and loss of product.

15. (a) Because cordycepin lacks a 3′-OH group, it cannot participate in 3′ → 5′ bond formation. (b) Because the poly(A) tail is a long stretch of adenosine nucleotides, the likelihood that a molecule of cordycepin would become incorporated is higher than with most RNA. (c) Yes, it must be converted into cordycepin 5′-triphosphate.
16. There are $2^8 = 256$ possible products.
17. The relationship between the −10 and −35 sequences could be affected by torsional strain. Since topoisomerase II introduces negative supercoils in DNA, this prevents this enzyme from overstimulating the expression of its own gene.
18. Ser-Ile-Phe-His-Pro-Stop
19. A mutation that disrupted the normal AAUAAA recognition sequence for the endonuclease could account for this finding. In fact, a change from U to C in this sequence caused this defect in a thalassemic patient. Cleavage occurred at the AAUAAA 900 nucleotides downstream from this mutant AACAAA site.
20. One possibility is that the 3′ end of the poly(U) donor strand cleaves the phosphodiester bond on the 5′ side of the insertion site. The newly formed 3′ terminus of the acceptor strand then cleaves the poly(U) strand on the 5′ side of the nucleotide that initiated the attack. In other words, a uridine residue could be added by two transesterification reactions. This postulated mechanism is similar to the one in RNA splicing.
21. Alternative splicing, RNA editing. Covalent modification of the proteins subsequent to synthesis.
22. Attach an oligo(dT) or oligo(U) sequence to an inert support to create an affinity column. When RNA is passed through the column, only poly(A)-containing RNA will be retained.
23. (a) Different amounts of RNA are present for the various genes.
(b) Although all of the tissues have the same genes, the genes are expressed to different extents in different tissues.
(c) These genes are called housekeeping genes—genes that most tissues express. They might include genes for glycolysis or citric acid cycle enzymes.
(d) The point of the experiment is to determine which genes are initiated in vivo. The initiation inhibitor is added to prevent initiation at start sites that may have been activated during the isolation of the nuclei.
24. DNA is the single strand that forms the trunk of the tree. Strands of increasing length are RNA molecules; the beginning of transcription is where growing chains are the smallest; the end of transcription is where chain growth stops. Direction is left to right. Many enzymes are actively transcribing each gene.

Chapter 30

1. The Oxford English Dictionary defines translation as the action or process of turning from one language into another. Protein synthesis converts nucleic acid sequence information into amino acid sequence information.
2. An error frequency of 1 incorrect amino acid every 10^4 incorporations allows for the rapid and accurate synthesis of proteins as large as 1000 amino acids. Higher error rates would result in too many defective proteins. Lower error rates would likely slow the rate of protein synthesis without a significant gain in accuracy.
3. (i) Each is a single chain. (ii) They contain unusual bases. (iii) Approximately half of the bases are base-paired to form double helices. (iv) The 5′ end is phosphorylated and is usually pG. (v) The amino acid is attached to the hydroxyl group of the A residue of the CCA sequence at the 3′ end of the tRNA. (vi) The anticodon is located in a loop near the center of the tRNA sequence. (vii) The molecules are L-shaped.
4. First is the formation of the aminoacyl adenylate, which then reacts with the tRNA to form the aminoacyl-tRNA. Both steps are catalyzed by aminoacyl-tRNA synthetase.
5. Unique features are required so that the aminoacyl-tRNA synthetases can distinguish among the tRNAs and attach the correct amino acid to the proper tRNA. Common features are required because all tRNAs must interact with the same protein-synthesizing machinery.
6. An activated amino acid is one linked to the appropriate tRNA.
7. (a) No; (b) no; (c) yes.
8. The ATP is cleaved to AMP and PP_i. Consequently, a second ATP is required to convert AMP into ADP, the substrate for oxidative phosphorylation.
9. Amino acids larger than the correct amino acid cannot fit into the active site of the tRNA. Smaller but incorrect amino acids that become attached to the tRNA fit into the editing site and are cleaved from the tRNA.
10. Recognition sites on both faces of the tRNAs may be required to uniquely identify the 20 different tRNAs.
11. The first two bases in a codon form Watson–Crick base pairs that are checked for fidelity by bases of the 16S rRNA. The third base is not inspected for accuracy, and so some variation is tolerated.
12. Four bands: light, heavy, a hybrid of light 30S and heavy 50S, and a hybrid of heavy 30S and light 50S.
13. Two hundred molecules of ATP are converted into 200 AMP + 400 P_i to activate the 200 amino acids, which is equivalent to 400 molecules of ATP. One molecule of GTP is required for initiation, and 398 molecules of GTP are needed to form 199 peptide bonds.
14. The reading frame is a set of contiguous, nonoverlapping three-nucleotide codons that begins with a start codon and ends with a stop codon.
15. A mutation caused by the insertion of an extra base can be suppressed by a tRNA that contains a fourth base in its anticodon. For example, UUUC rather than UUU is read as the codon for phenylalanine by a tRNA that contains 3′-AAAG-5′ as its anticodon.
16. One approach is to synthesize a tRNA that is acylated with a reactive amino acid analog. For example, bromoacetyl-phenylalanyl-tRNA is an affinity-labeling reagent for the P site of *E. coli* ribosomes.
17. The sequence GAGGU is complementary to a sequence of five bases at the 3′ end of 16S rRNA and is located several bases upstream of an AUG start codon. Hence, this region is a start signal for protein synthesis. The replacement of G by A would be expected to weaken the interaction of this mRNA with the 16S rRNA and thereby diminish its effectiveness as an initiation signal. In fact, this mutation results in a 10-fold decrease in the rate of synthesis of the protein specified by this mRNA.
18. The peptide would be Phe-Cys-His-Val-Ala-Ala. The codons UGC and UGU encode cysteine but, because the cysteine was modified to alanine, alanine is incorporated in place of cysteine.
19. Proteins are synthesized from the amino to the carboxyl end on ribosomes, whereas they are synthesized in the reverse direction in the solid-phase method. The activated intermediate in ribosomal synthesis is an aminoacyl-tRNA; in the solid-phase method, it is the adduct of the amino acid and dicyclohexylcarbodiimide.
20. GTP is not hydrolyzed until aminoacyl-tRNA is delivered to the A site of the ribosome. An earlier hydrolysis of GTP would be wasteful because EF-Tu–GDP has little affinity for aminoacyl-tRNA.
21. The translation of an mRNA molecule can be blocked by antisense RNA, an RNA molecule with the complementary sequence. The antisense–sense RNA duplex cannot serve as a template for translation; single-stranded mRNA is required. Furthermore, the antisense–sense duplex is degraded by nucleases. Antisense RNA added to the external medium is spontaneously taken up by

many cells. A precise quantity can be delivered by microinjection. Alternatively, a plasmid encoding the antisense RNA can be introduced into target cells.

22. (a) A_5. (b) $A_5 > A_4 > A_3 > A_2$. (c) Synthesis is from the amino terminus to the carboxyl terminus.

23. These enzymes convert nucleic acid information into protein information by interpreting the tRNA and linking it to the proper amino acid.

24. The rate would fall because the elongation step requires that the GTP be hydrolyzed before any further elongation can take place.

25. Protein factors modulate the initiation of protein synthesis. The role of IF1 and IF3 is to prevent premature binding of the 30S and 50S ribosomal subunits, whereas IF2 delivers Met-$tRNA_f$ to the ribosome. Protein factors are also required for elongation (EF-G and EF-Tu), for termination (release factors, RFs), and for ribosome dissociation (ribosome release factors, RRFs).

26. The signal sequence, signal-recognition particle (SRP), the SRP receptor, and the translocon.

27. The formation of peptide bonds, which in turn are powered by the hydrolysis of the aminoacyl-tRNAs.

28. The Shine–Dalgarno sequence of the mRNA base-pairs with a part of the 16S rRNA of the 30S subunit, which positions the subunit so that the initiator AUG is recognized.

29.

	Bacteria	**Eukaryote**
Ribosome size	70S	80S
mRNA	Polycistronic	Not polycistronic
Initiation	Shine–Dalgarno is required	First AUG is used
Protein factors	Required	Many more required
Relation to transcription	Translation can start before transcription is completed	Transcription and translation are spatially separated
First amino acid	fMet	Met

30. The SRP binds to the signal sequence and inhibits further translation. The SRP ushers the inhibited ribosome to the ER, where it interacts with the SRP receptor (SR). The SRP–SR complex binds the translocon and simultaneously hydrolyzes GTP. On GTP hydrolysis, SRP and SR dissociate from each other and from the ribosome. Protein synthesis resumes and the nascent protein is channeled through the translocon.

31. The alternative would be to have a single ribosome translating a single mRNA molecule. The use of polysomes allows more protein synthesis per mRNA molecule in a given period of time and thus the production of more protein.

32. (a) 1, 2, 3, 5, 6, 10; (b) 1, 2, 7, 8; (c) 1, 4, 8, 9.

33. Transfer RNAs have roles in several recognition processes. A tRNA must be recognized by the appropriate aminoacyl-tRNA synthetase, and the tRNA must interact with the ribosome and, in particular, with the peptidyl transferase.

34. The aminoacyl-tRNA can be initially synthesized. However, the side-chain amino group attacks the ester linkage to form a six-membered amide, releasing the tRNA.

35. (a, d, and e) Type 2; (b, c, and f) Type 1.

36. The error rates of DNA, RNA, and protein synthesis are of the order of 10^{-10}, 10^{-5}, and 10^{-4}, respectively, per nucleotide (or amino acid) incorporated. The fidelity of all three processes depends on the precision of base-pairing to the DNA or mRNA template. Few errors are corrected in RNA synthesis. In contrast, the fidelity of DNA synthesis is markedly increased by the $3' \rightarrow 5'$ proofreading nuclease activity and by postreplicative repair. In protein synthesis, the mischarging of some tRNAs is corrected by the hydrolytic action of aminoacyl-tRNA synthetase. Proofreading also takes place when aminoacyl-tRNA occupies the A site on the ribosome; the GTPase activity of EF-Tu sets the pace of this final stage of editing.

37. EF-Ts catalyzes the exchange of GTP for GDP bound to EF-Tu. In G-protein cascades, an activated 7TM receptor catalyzes GTP–GDP exchange in a G protein.

38. The α subunits of G proteins are inhibited by a similar mechanism in cholera and whooping cough (Section 14.5).

39. Glu-$tRNA^{Gln}$ is formed by misacylation. The activated glutamate is subsequently amidated to form Gln-$tRNA^{Gln}$. Ways in which glutamine is formed from glutamate were discussed in Section 24.2. In regard to *H. pylori,* a specific enzyme, Glu-$tRNA^{Gln}$ amidotransferase, catalyzes the following reaction:

$$\text{Gln} + \text{Glu-tRNA}^{\text{Gln}} + \text{ATP} \longrightarrow \text{Gln-tRNA}^{\text{Gln}} + \text{Glu} + \text{ADP} + P_i$$

Glu-$tRNA^{Glu}$ is not a substrate for the enzyme; so the transferase must also recognize aspects of the structure of $tRNA^{Gln}$.

40. The primary structure determines the three-dimensional structure of the protein. Thus, the final phase of information transfer from DNA to RNA to protein synthesis is the folding of the protein into its functional state.

41. (a) eIF-4H has two effects: (1) the extent of unwinding is increased and (2) the rate of unwinding is increased, as indicated by the increased rise in activity at early reaction times.
(b) To firmly establish that the effect of eIF-H4 was not due to any inherent helicase activity.
(c) Half-maximal activity was achieved at 0.11 μM of eIF-4H. Therefore, maximal stimulation would be achieved at a ratio of 1:1.
(d) eIF-4H enhances the rate of unwinding of all helices, but the effect is greater as the helices increase in stability.
(e) The results in graph C suggest that eIF-4H increases the processivity.

42. (a) The three peaks represent, from left to right, the 40S ribosomal subunit, the 60S ribosomal subunit, and the 80S ribosome.
(b) Not only are ribosomal subunits and the 80S ribosome present, but polysomes of various lengths also are apparent. The individual peaks in the polysome region represent polysomes of discrete length.
(c) The treatment significantly inhibited the number of polysomes while increasing the number of free ribosomal subunits. This outcome could be due to inhibited protein-synthesis initiation or inhibited transcription.

Chapter 31

1. (a) Cells will express β-galactosidase, *lac* permease, and thiogalactoside transacetylase even in the absence of lactose. (b) Cells will express β-galactosidase, *lac* permease, and thiogalactoside transacetylase even in the absence of lactose. (c) The levels of catabolic enzymes such as β-galactosidase and arabinose isomerase will remain low even at low levels of glucose.

2. The concentration is $1/(6 \times 10^{23})$ moles per 10^{-15} liter $= 1.7 \times 10^{-9}$ M. Because $K_d = 10^{-13}$ M, the single molecule should be bound to its specific binding site.

3. The number of possible 8-bp sites is $4^8 = 65{,}536$. In a genome of 4.6×10^6 base pairs, the average site should appear $4.6 \times 10^6/65{,}536 = 70$ times. Each 10-bp site should appear 4 times. Each 12-bp site should appear 0.27 times (many 12-bp sites will not appear at all).

4. The *lac* repressor does not bind DNA when the repressor is bound to a small molecule (the inducer), whereas the *pur* repressor binds

DNA only when the repressor is bound to a small molecule (the corepressor). The *E. coli* genome contains only a single *lac* repressor-binding region, whereas it has many sites for the *pur* repressor.
5. Anti-inducers bind to the conformation of repressors, such as the *lac* repressor, that are capable of binding DNA. They occupy a site that overlaps that for the inducer and, therefore, compete for binding to the repressor.
6. The inverted repeat may be a binding site for a dimeric DNA-binding protein or it may correspond to a stem-loop structure in the encoded RNA.
7. Bacteriophage λ would be more likely to enter the lytic phase because the cooperative binding of the λ repressor to O_R2 and O_R1, which supports the lysogenic pathway, would be disrupted.
8. λ repressor gene −10 region GATTTA −35 region TAGATA
Cro gene −10 region TAATGG −35 region TTGACT
There are four differences in the −10 region and three differences in the −35 region.
9. Increased Cro concentration reduces the expression of the λ repressor gene. Increased λ repressor concentration reduces the expression of the Cro gene. At low λ repressor concentrations, increased λ repressor concentration increases the expression of the λ repressor gene. At higher λ repressor concentrations, increased λ repressor concentration decreases the expression of the λ repressor gene.
10. Normally, bacterial mRNAs have a leader sequence in which a Shine–Dalgarno sequence precedes the AUG start codon. The absence of a leader would be expected to lead to inefficient translation.
11. Add each compound to a culture of *V. fischeri* at low density and look for the development of luminescence.
12. ACC, 7; ACA, 1; ACU, 0; ACG, 0.
13. Retention. The reaction likely proceeds in two steps: Attack of Glu 537 on the carbon of the galactose moiety (with inversion) followed by attack of water on this carbon (with a second inversion) to release the galactose from the enzyme.
14. The footprint appears to have approximately 50% of its intensity near 3.7 nM so that the dissociation constant is approximately 3.7 nM, corresponding to a standard free energy of binding of −48 kJ/mol (−11 kcal/mol) at T = 298K.

Chapter 32

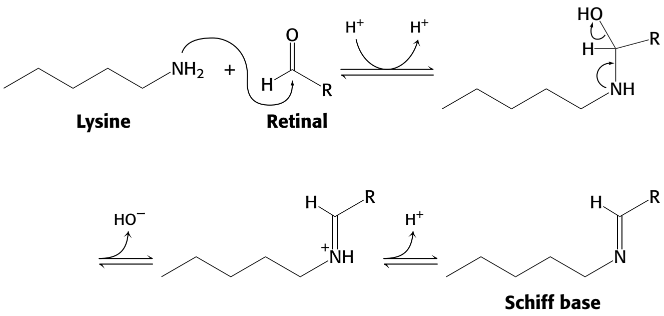
1. The distribution of charged amino acids is H2A (13 K, 13 R, 2 D, 7 E, charge = +17), H2B (20 K, 8 R, 3 D, 7 E, charge = +18), H3 (13 K, 18 R, 4 D, 7 E, charge = +20), H4 (11 K, 14 R, 3 D, 4 E, charge = +18). The total charge of the histone octamer is estimated to be $2 \times (17 + 18 + 20 + 18) = +146$. The total charge on 150 base pairs of DNA is −300. Thus, the histone octamer neutralizes approximately one-half of the charge.
2. The presence of a particular DNA fragment could be detected by hybridization, by PCR, or by direct sequencing.
3. The total length of the DNA is estimated to be $145 \text{ bp} \times 3.4 \text{ Å/bp} = 493 \text{ Å}$, which represents 1.75 turns or $1.75 \times 2\pi r = 11.0r$. Thus, the radius is estimated to be $r = 493 \text{ Å}/11.0 = 44.8 \text{ Å}$.
4. 5-Azacytidine cannot be methylated. Some genes, normally repressed by methylation, will be active. 5-Azacytidine is also an inhibitor of DNA methyltransferase, resulting in lower levels of methylation and, hence, lower levels of gene repression.
5. Proteins containing these domains will be targeted to methylated DNA in repressed promoter regions. They would likely bind in the major groove because that is where the methyl group is located.
6. Gene expression is not expected to respond to the presence of estrogen. However, genes for which expression normally responds to estrogen will respond to the presence of progesterone.
7. The acetylation of lysine will reduce the charge from +1 to 0. The methylation of lysine will not reduce the charge.
8. On the basis of the pattern of cysteine and histidine residues, this region appears to contain three zinc-finger domains.
9. $10/4000 = 0.25\%$. 0.25% of 12 Mb = 30 kilobase pairs.
10. Monomethylated-activated; Trimethylated-repressed.
11. The addition of an IRE to the 5′ end of the mRNA is expected to block translation in the absence of iron. The addition of an IRE to the 3′ end of the mRNA is not expected to block translation, but it might affect mRNA stability.
12. The sequences of all of the mRNAs would be searched for sequences that are fully or nearly complementary to the sequence of the miRNA. These sequences would be candidates for regulation by this mRNA.
13. The amino group of the lysine residue, formed from the protonated form by a base, attacks the carbonyl group of acetyl CoA to generate a tetrahedral intermediate. This intermediate collapses to form the amide bond and release CoA.
14. In mouse DNA, most of the HpaII sites are methylated and therefore not cut by the enzyme, resulting in large fragments. Some small fragments are produced from CpG islands that are unmethylated. For *Drosophila* and *E. coli* DNA, there is no methylation and all sites are cut.

Chapter 33

1. The transgenic nematode would avoid the compound. The identity of the ligand is determined by the receptor, whereas the behavioral response is dictated by the neuron in which the receptor is expressed.
2. Only a mixture of compounds C_5-COOH and HOOC-C_7-COOH is predicted to yield this pattern.
3. Bitter and sweet sensations are mediated by G proteins coupled to 7TM receptors, leading to millisecond time resolution. Salty and sour sensations are mediated directly by ion channels, which may lead to faster time resolution.
4. Sound travels 0.15 m in 428 μs. The human hearing system is capable of sensing time differences of close to a microsecond, and so the difference in arrival times at the two ears is substantial. A system based on G proteins is unlikely to be able to reliably distinguish between signals arriving at the two ears, because G proteins typically respond in milliseconds.
5. If a plant tastes bitter, animals will avoid eating it even if it is nontoxic, which may provide a selective advantage to the plant.
6. Using mice in which either the gene for T1R1 or the gene for T1R3 has been disrupted, test the taste responses of these mice to glutamate, aspartate, and a wide variety of other amino acids.
7. 530 nm light will be absorbed to some degree by all three photoreceptors, with the largest absorption from the "green" receptor.
8. These women have four functional color receptors: blue, red, green, and a red–green hybrid. The additional color receptor allows some colors that appear identical to most people to be distinguished.
9. 380 (one for each receptor); there are $(380 \times 379)/2! = 72{,}010$ combinations of two receptors; $(380 \times 379 \times 378)/3! = 9{,}073{,}260$ combinations of three receptors.
10. The absorption of light converts 11-*cis*-retinal into all-*trans*-retinal.
11. These compounds are enantiomers and must bind to protein receptors to elicit a smell. Even these subtle structural differences can affect relative receptor binding affinities and, hence, the elicited odor.
12. Vision: cGMP-gated channel; taste: amiloride-sensitive sodium channel; hearing; tip-link channel.

13. For all senses, ATP hydrolysis is required to generate and maintain ion gradients and membrane potential. Olfaction: ATP is required for the synthesis of cAMP. Gustation: ATP is required for the synthesis of cyclic nucleotides, and GTP is required for the action of gustducin in the detection of bitter and sweet tastes. Vision: GTP is required for the synthesis of cGMP and for the action of transducin. Hearing and touch: ATP hydrolysis is required to generate and maintain ion gradients and membrane potential and may be required for other roles as well.

14.



Chapter 34

1. The innate immune system responds rapidly to common features present in many pathogens. The genes for the innate immune system's key molecules are expressed without substantial modification. In contrast, the adaptive immune system responds to specific features present only in a given pathogen. Its genes undergo significant rearrangement and mutation to enable specific recognition of a vast number of potential binding surfaces.

2. VJ and V(D)J recombination; variability in segment joining by the action of terminal deoxyribonucleotidyl transferase; somatic mutation.

3. *Affinity* refers to the strength of a single interaction; *avidity* refers to the cumulative strength of multiple independent binding interactions. Avidity may play a significant role in the interaction between IgM and antigen because this immunoglobulin class features 10 binding sites.

4. The intracellular signaling domain common to each of the TLRs is responsible for docking other proteins and reporting that a targeted pathogen-associated molecular pattern (PAMP), such as LPS, has been detected. If a mutation within this domain interfered with the intracellular docking and signal transduction, then TLR4 would not respond to LPS.

5. Viruses that contain dsRNA genomes would be expected to stimulate a TLR3-mediated immune response.

6. (a) $\Delta G^{\circ\prime} = -37\ \mathrm{kJ\ mol^{-1}}(-8.9\ \mathrm{kcal\ mol^{-1}})$

(b) $K_a = 3.3 \times 10^6\ \mathrm{M^{-1}}$

(c) $k_{on} = 4 \times 10^8\ \mathrm{M^{-1}\ s^{-1}}$. This value is close to the diffusion-controlled limit for the combination of a small molecule with a protein (p. 231). Hence, the extent of structural change is likely to be small; extensive conformational transitions take time.

7. The fluorescence enhancement and the shift to blue indicate that water is largely excluded from the binding site when the hapten is bound. Hydrophobic interactions contribute significantly to the formation of most antigen–antibody complexes.

8. (a) An antibody combining site is formed by CDRs from both the H and the L chains. The V_H and V_L domains are essential. A small proportion of F_{ab} fragments can be further digested to produce F_v, a fragment that contains just these two domains. C_H1 and C_L contribute to the stability of F_{ab} but not to antigen binding.

(b) A synthetic F_v analog 248 residues long was prepared by expressing a synthetic gene consisting of a V_H gene joined to a V_L gene through a linker. [J. S. Huston et al., *Proc. Natl. Acad. Sci. U.S.A.* 85:5879–5883, 1988.]

9. (a) Multivalent antigens lead to the dimerization or oligomerization of transmembrane immunoglobulins, an essential step in their activation. This mode of activation is reminiscent of that of receptor tyrosine kinases (Section 14.2).

(b) An antibody specific for a transmembrane immunoglobulin will activate a B cell by cross-linking these receptors. This experiment can be carried out by using, for example, a goat antibody to cross-link receptors on a mouse B cell.

10. B cells do not express T-cell receptors. The hybridization of T-cell cDNAs with B-cell mRNAs removes cDNAs that are expressed in both cells. Hence, the mixture of cDNAs subsequent to this hybridization are enriched in those encoding T-cell receptors. This procedure, called subtractive hybridization, is generally useful in isolating low-abundance cDNAs. Hybridization should be carried out by using mRNAs from a closely related cell that does not express the gene of interest. [See S. M. Hedrick, D. I. Cohen, E. A. Nielsen, and M. M. Davis, *Nature* 308:149–153, 1984, for an interesting account of how this method was used to obtain genes for T-cell receptors.]

11. TLR4 is the receptor for LPS, a toxin found specifically in the walls of Gram-negative bacteria. Mutations that inhibit the function of TLR4 impair an affected person's defenses against this class of bacteria.

12. If the HLA alleles are not matched, then the recipient's T cell receptors will identify the MHC proteins of the transplanted tissue as nonself and transplant rejection is likely.

13. Purify an antibody with a specificity to one antigen. Unfold the antibody and allow it to re-fold either in the presence of the antigen or in the absence of the antigen. Test the re-folded antibodies for antigen-binding ability.

14. In some cases, V–D–J rearrangement will result in combining V, D, and J segments out of frame. mRNA molecules produced from such rearranged genes will produce truncated molecules if translated. This possibility is excluded by degrading the mRNA.

15. The mutant bacteria may still stimulate an immune response without causing disease. Hence, they may be valuable starting points for the design of a live attenuated vaccine for the original pathogenic strain.

16. The peptide is LLQATYSAV (L in second position, V in last).

17. Catalysis is likely to require a base for removing a proton from a water molecule. A histidine, glutamate, or aspartate residue is most likely. In addition, a potential hydrogen-bond donor may be present and will interact with the negatively charged oxygen atom that forms in the transition state.

18. A phosphotyrosine residue in the carboxyl terminus of Src and related protein tyrosine kinases binds to its own SH2 domain to generate the inhibited form of Src (Section 14.5). Removal of the phosphoryl group from this residue will activate the kinase.

19. (a) $K_d = 10^{-7}$ M; (b) $K_d = 10^{-9}$ M. The gene was probably generated by a point mutation in the gene for antibody A rather than by de novo rearrangement.

Chapter 35

1. (a) Skeletal muscle and eukaryotic cilia derive their free energy from ATP hydrolysis; the bacterial flagellar motor uses a proton-motive force.

(b) Skeletal muscle requires myosin and actin. Eukaryotic cilia require microtubules and dynein. The bacterial flagellar motor requires MotA, MotB, and FliG, as well as many ancillary components.

2. 6400 Å/80 Å = 80 body lengths per second. For a 10-foot automobile, this body-length speed corresponds to a speed of 80×10 feet = 800 feet per second, or 545 miles per hour.

3. 4 pN = 8.8×10^{-13} pounds. The weight of a single motor domain is 100,000 g $\text{mol}^{-1}/(6.023 \times 10^{23}$ molecules $\text{mol}^{-1}) = 1.7 \times 10^{-19}$ g = 3.7×10^{-22} pounds. Thus, a motor domain can lift $(8.8 \times 10^{-13}/3.7 \times 10^{-22}) = 2.4 \times 10^{9}$ times its weight.

4. Both actin filaments and microtubules are built from subunits and these subunits bind and hydrolyze nucleoside triphosphates. Actin filaments are built of a single type of subunit and these subunits bind ATP. Microtubules are built of two different types of subunits and these subunits bind GTP.

5. The light chains in myosin stiffen the lever arm. The light chains in kinesin bind cargo to be transported.

6. After death, the ratio of ADP to ATP increases rapidly. In the ADP form, myosin motor domains bind tightly to actin. Myosin–actin interactions are possible because the drop in ATP concentration also allows the calcium concentration to rise, clearing the blockage of actin by tropomyosin through the action of the troponin complex.

7. Above its critical concentration, ATP-actin will polymerize. The ATP will hydrolyze through time to form ADP-actin, which has a higher critical concentration. Thus, if the initial subunit concentration is between the critical concentrations of ATP-actin and ADP-actin, filaments will form initially and then disappear on ATP hydrolysis.

8. A one-base step is approximately 3.4 Å = 3.4×10^{-4} μm. If a stoichiometry of one molecule of ATP per step is assumed, this distance corresponds to a velocity of 0.017 μm s^{-1}. Kinesin moves at a velocity of 6400 Å per second, or 0.64 μm s^{-1}.

9. A proton-motive force across the plasma membrane is necessary to drive the flagellar motor. Under conditions of starvation, this proton-motive force is depleted. In acidic solution, the pH difference across the membrane is sufficient to power the motor.

10. The mean distance between tumbles would be longer when the bacterium is moving up a gradient of a chemoattractant.

11. (a) 1.13×10^{-9} dyne
(b) 6.8×10^{14} erg
(c) 6.6×10^{-11} erg per 80 molecules of ATP. A single kinesin motor provides more than enough free energy to power the transport of micrometer-size cargoes at micrometer-per-second velocities.

12. The spacing between identical subunits on microtubules is 8 nm. Thus, a kinesin molecule with a step size that is not a multiple of 8 nm would have to be able to bind at more than one type of site on the microtubule surface.

13. KIF1A must be tethered to an additional microtubule-binding element that retains an attachment to the microtubule when the motor domain releases.

14. Filaments built from subunits can be arbitrarily long, can be dynamically assembled and disassembled, and require only a small amount of genetic information to encode.

15. Protons still flow from outside to inside the cell. Each proton might pass into the outer half-channel of one MotA–MotB complex, bind to the MS ring, rotate clockwise, and pass into the inner half-channel of the neighboring MotA–MotB complex.

16. At a high concentration of calcium ion, Ca^{2+} binds to calmodulin. In turn, calmodulin binds to a protein kinase that phosphorylates myosin light chains and activates it. At low calcium ion concentration, the light chains are dephosphorylated by a Ca^{2+}-independent phosphatase.

17. (a) The value of k_{cat} is approximately 13 molecules per second, whereas the K_{M} value for ATP is approximately 12 μM.
(b) The step size is approximately $(380 - 120)/7 = 37$ nm.
(c) The step size is very large, which is consistent with the presence of six light-chain-binding sites and, hence, very long lever arms. The rate of ADP release is essentially identical with the overall k_{cat}; so ADP release is rate limiting, which suggests that both motor domains can bind to sites 37 nm apart simultaneously. ADP release from the hindmost domain allows ATP to bind, leading to actin release and lever-arm motion.

Chapter 36

1. (a) Before; (b) after; (c) after; (d) after; (e) before; (f) after.

2. (a) Yes; (b) yes; (c) no (MW > 600).

3. If computer programs could estimate log(P) values on the basis of chemical structure, then the required laboratory time for drug development could be shortened. The determination of the relative solubilities of pharmaceutical candidates by allowing each compound to equilibrate between water and an organic phase would no longer be necessary.

4. Perhaps *N*-acetylcysteine would conjugate to some of the *N*-acetyl-*p*-benzoquinone imine that is produced by the metabolism of acetaminophen, thereby preventing the depletion of the liver's supply of glutathione.

5. In Phase I clinical trials, approximately 10 to 100 usually healthy volunteers are typically enrolled in a study designed to assess safety. In contrast, a larger number of subjects are enrolled in a typical Phase II trial. Moreover, these persons may benefit from the drug administered. In a Phase II trial, efficacy, dosage, and safety can be assessed.

6. The binding of other drugs to albumin could cause extra coumadin to be released. (Albumin is a general carrier for hydrophobic molecules.)

7. A drug that inhibits a P450 enzyme may dramatically affect the disposition of another drug that is metabolized by that same enzyme. If this inhibited metabolism is not accounted for when dosing, the second drug may reach very high, and sometimes toxic, levels in the blood.

8. Unlike competitive inhibition, noncompetitive inhibition cannot be overcome with additional substrate. Hence, a drug that acts by a noncompetitive mechanism will be unaffected by changing levels of the physiological substrate.

9. An inhibitor of MDR could prevent the efflux of a chemotherapeutic drug from tumor cells. Hence, this type of an inhibitor could be useful in averting resistance to cancer chemotherapy.

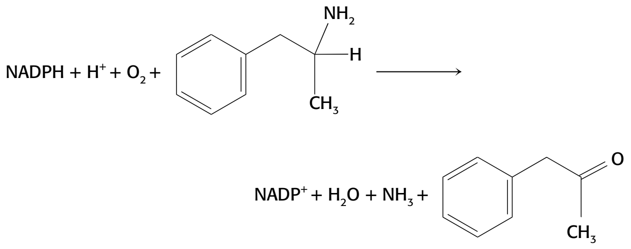
10. Agents that inhibit one or more enzymes of the glycolytic pathway could act to deprive trypanosomes of energy and thus be useful for treating sleeping sickness. A difficulty is that glycolysis in the host cells also would be inhibited.

11. Imatinib is an inhibitor of the Bcr-Abl kinase, a mutant kinase present only in tumor cells that have undergone a translocation between chromosomes 9 and 22 (Figure 14.34). Before initiating treatment with imatinib, we could sequence the DNA of the tumor cells and determine (a) whether this translocation has taken place and (b) whether the sequence of *bcr-abl* carries any mutations that would render the kinase resistant

to imatinib. If the translocation has not taken place or if the gene carries resistance mutations, then imatinib would likely not be an effective treatment for the patient carrying this particular tumor.

12. Sildenafil increases cGMP levels by inhibiting the phosphodiesterase-mediated breakdown of cGMP to GMP. Intracellular cGMP levels can also be increased by activating its synthesis. This activation can be achieved with the use of NO donors (such as sodium nitroprusside and nitroglycerin) or compounds that activate guanylate cyclase activity. Drugs that act by the latter mechanism are currently in clinical trials.

13. A reasonable mechanism would be an oxidative deamination following an overall mechanism similar to that in Figure 36.10, with release of ammonia.



14. $K_I \approx 0.3$ nM. $IC_{50} \approx 2.0$ nM. Yes, compound A should be effective when taken orally because 400 nM is much greater than the estimated values of K_I and IC_{50}.

SELECTED READINGS

Chapter 2

Where to Start

Service, R. F. 2008. Problem solved* (*sort of) (a brief review of protein folding). *Science* 321:784–786.

Doolittle, R. F. 1985. Proteins. *Sci. Am.* 253(4):88–99.

Richards, F. M. 1991. The protein folding problem. *Sci. Am.* 264(1):54–57.

Weber, A. L., and Miller, S. L. 1981. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.* 17:273–284.

Books

Petsko, G. A., and Ringe, D. 2004. *Protein Structure and Function*. New Science Press.

Tanford, C., and Reynolds, J. 2004. *Nature's Robots: A History of Proteins.* Oxford.

Branden, C., and Tooze, J. 1999. *Introduction to Protein Structure* (2d ed.). Garland.

Creighton, T. E. 1992. *Proteins: Structures and Molecular Principles* (2d ed.). W. H. Freeman and Company.

Conformation of Proteins

Smock, R. G., and Gierasch, L. M. 2009. Sending signals dynamically. *Science* 324:198–203.

Tokuriki, N., and Tawfik, D. S. 2009. Protein dynamism and evolvability. *Science* 324:203–207.

Pace, C. N., Grimsley, G. R., and Scholtz, J. M. 2009. Protein ionizable groups: p*K* values and their contribution to protein stability and solubility. *J. Biol. Chem.* 284:13285–13289.

Breslow, R., and Cheng, Z.-L. 2009. On the origin of terrestrial homochirality for nucleosides and amino acids. *Proc. Natl. Acad. Sci. U.S.A.* 106:9144–9146.

Secondary Structure

Shoulders, M. D., and Raines, R. T. 2009. Collagen structure and stability. *Annu. Rev. Biochem.* 78:929–958.

O'Neil, K. T., and DeGrado, W. F. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651.

Zhang, C., and Kim, S. H. 2000. The anatomy of protein beta-sheet topology. *J. Mol. Biol.* 299:1075–1089.

Regan, L. 1994. Protein structure: Born to be beta. *Curr. Biol.* 4:656–658.

Srinivasan, R., and Rose, G. D. 1999. A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 96:14258–14263.

Intrinsically Unstructured Proteins

Galea, C. A., Wang, Y., Sivakolundu, S. G., and Kriwacki, R. W. 2008. Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. *Biochemistry* 47:7598–7609.

Raychaudhuri, S., Dey, S., Bhattacharyya, N. P., and Mukhopadhyay, D. 2009. The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS One* 4:e5566.

Tompa, P., and Fuxreiter, M. 2008. Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* 33:2–8.

Tuinstra, R. L., Peterson, F. C., Kutlesa, E. S., Elgin, S., Kron, M. A., and Volkman, B. F. 2008. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl. Acad. Sci. U.S.A.* 105:5057–5062.

Domains

Jin, J., Xie, X., Chen, C., Park, J. G., Stark, C., James, D. A., Olhovsky, M., Lindinger, R., Mao, Y., and Pawson, T. 2009. Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.* 2:ra76.

Bennett, M. J., Choe, S., and Eisenberg, D. 1994. Domain swapping: Entangling alliances between proteins. *Proc. Natl. Acad. Sci. U.S.A.* 91:3127–3131.

Bergdoll, M., Eltis, L. D., Cameron, A. D., Dumas, P., and Bolin, J. T. 1998. All in the family: Structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly. *Protein Sci.* 7:1661–1670.

Hopfner, K. P., Kopetzki, E., Kresse, G. B., Bode, W., Huber, R., and Engh, R. A. 1998. New enzyme lineages by subdomain shuffling. *Proc. Natl. Acad. Sci. U.S.A.* 95:9813–9818.

Ponting, C. P., Schultz, J., Copley, R. R., Andrade, M. A., and Bork, P. 2000. Evolution of domain families. *Adv. Protein Chem.* 54:185–244.

Protein Folding

Caughey, B., Baron, G. S., Chesebro, B., and Jeffrey, M. 2009. Getting a grip on prions: Oligomers, amyloids, and pathological membrane interactions. *Annu. Rev. Biochem.* 78:177–204.

Cobb, N. J., and Surewicz, W. K. 2009. Prion diseases and their biochemical mechanisms. *Biochemistry* 48:2574–2585.

Soto, C. 2011. Prion diseases: The end of the controversy? *Trends Biochem. Sci.* 36:151–158.

Daggett, V., and Fersht, A. R. 2003. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28:18–25.

Selkoe, D. J. 2003. Folding proteins in fatal ways. *Nature* 426:900–904.

Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science* 181:223–230.

Baldwin, R. L., and Rose, G. D. 1999. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* 24:26–33.

Baldwin, R. L., and Rose, G. D. 1999. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* 24:77–83.

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. 2003. Design of a novel globular protein with atomic-level accuracy. *Science* 302:1364–1368.

Staley, J. P., and Kim, P. S. 1990. Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. *Nature* 344:685–688.

Covalent Modification of Proteins

Tarrant, M. K., and Cole, P. A. 2009. The chemical biology of protein phosphorylation. *Annu. Rev. Biochem.* 78:797–825.

Krishna, R. G., and Wold, F. 1993. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas. Mol. Biol.* 67:265–298.

Aletta, J. M., Cimato, T. R., and Ettinger, M. J. 1998. Protein methylation: A signal event in post-translational modification. *Trends Biochem. Sci.* 23:89–91.

Tsien, R. Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem.* 67:509–544.

Chapter 3

Where to Start

Sanger, F. 1988. Sequences, sequences, sequences. *Annu. Rev. Biochem.* 57:1–28.

Merrifield, B. 1986. Solid phase synthesis. *Science* 232:341–347.

Hunkapiller, M. W., and Hood, L. E. 1983. Protein sequence analysis: Automated microsequencing. *Science* 219:650–659.

Milstein, C. 1980. Monoclonal antibodies. *Sci. Am.* 243(4):66–74.

Moore, S., and Stein, W. H. 1973. Chemical structures of pancreatic ribonuclease and deoxyribonuclease. *Science* 180:458–464.

Books

Methods in Enzymology. Academic Press.

Wilson, K., and Walker, J. (Eds.). 2010. *Principles and Techniques of Practical Biochemistry* (7th ed.). Cambridge University Press.

Van Holde, K. E., Johnson, W. C., and Ho, P.-S. 1998. *Principles of Physical Biochemistry.* Prentice Hall.

Wilkins, M. R., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. 1997. *Proteome Research: New Frontiers in Functional Genomics (Principles and Practice).* Springer Verlag.

Johnstone, R. A. W. 1996. *Mass Spectroscopy for Chemists and Biochemists* (2d ed.). Cambridge University Press.

Kyte, J. 1994. *Structure in Protein Chemistry.* Garland.

Creighton, T. E. 1993. *Proteins: Structure and Molecular Properties* (2d ed.). W. H. Freeman and Company.

Cantor, C. R., and Schimmel, P. R. 1980. *Biophysical Chemistry.* W. H. Freeman and Company.

Protein Purification and Analysis

Blackstock, W. P., and Weir, M. P. 1999. Proteomics: Quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17:121–127.

Deutscher, M. (Ed.). 1997. *Guide to Protein Purification.* Academic Press.

Dunn, M. J. 1997. Quantitative two-dimensional gel electrophoresis: From proteins to proteomes. *Biochem. Soc. Trans.* 25:248–254.

Scopes, R. K., and Cantor, C. 1994. *Protein Purification: Principles and Practice* (3d ed.). Springer Verlag.

Aebersold, R., Pipes, G. D., Wettenhall, R. E., Nika, H., and Hood, L. E. 1990. Covalent attachment of peptides for high sensitivity solid-phase sequence analysis. *Anal. Biochem.* 187:56–65.

Ultracentrifugation and Mass Spectrometry

Steen, H., and Mann, M. 2004. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5:699–711.

Glish, G. L., and Vachet, R. W. 2003. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov*. 2:140–150.

Li, L., Garden, R. W., and Sweedler, J. V. 2000. Single-cell MALDI: A new tool for direct peptide profiling. *Trends Biotechnol.* 18:151–160.

Yates, J. R., 3d. 1998. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* 33:1–19.

Pappin, D. J. 1997. Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol. Biol.* 64:165–173.

Schuster, T. M., and Laue, T. M. 1994. *Modern Analytical Ultracentrifugation.* Springer Verlag.

Arnott, D., Shabanowitz, J., and Hunt, D. F. 1993. Mass spectrometry of proteins and peptides: Sensitive and accurate mass measurement and sequence analysis. *Clin. Chem.* 39:2005–2010.

Chait, B. T., and Kent, S. B. H. 1992. Weighing naked proteins: Practical, high-accuracy mass measurement of peptides and proteins. *Science* 257:1885–1894.

Edmonds, C. G., Loo, J. A., Loo, R. R., Udseth, H. R., Barinaga, C. J., and Smith, R. D. 1991. Application of electrospray ionization mass spectrometry and tandem mass spectrometry in combination with capillary electrophoresis for biochemical investigations. *Biochem. Soc. Trans.* 19:943–947.

Jardine, I. 1990. Molecular weight analysis of proteins. *Methods Enzymol.* 193:441–455.

Proteomics

Yates, J. R., 3rd. 2004. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* 33:297–316.

Weston, A. D., and Hood, L. 2004. Systems biology, proteomics, and the future of health care: Toward predictive, preventative, and personalized medicine. *J. Proteome Res.* 3:179–196.

Pandey, A., and Mann, M. 2000. Proteomics to study genes and genomes. *Nature* 405:837–846.

Dutt, M. J., and Lee, K. H. 2000. Proteomic analysis. *Curr. Opin. Biotechnol.* 11:176–179.

Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. 2000. The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.* 148:635–651.

X-ray Crystallography and NMR Spectroscopy

Rhodes, G. 2006. *Crystallography Made Crystal Clear.* Elsevier/Academic Press.

Moffat, K. 2003. The frontiers of time-resolved macromolecular crystallography: Movies and chirped X-ray pulses. *Faraday Discuss.* 122:65–88.

Bax, A. 2003. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* 12:1–16.

Wery, J. P., and Schevitz, R. W. 1997. New trends in macromolecular x-ray crystallography. *Curr. Opin. Chem. Biol.* 1:365–369.

Glusker, J. P. 1994. X-ray crystallography of proteins. *Methods Biochem. Anal.* 37:1–72.

Clore, G. M., and Gronenborn, A. M. 1991. Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252:1390–1399.

Wüthrich, K. 1989. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* 243:45–50.

Wüthrich, K. 1986. *NMR of Proteins and Nucleic Acids.* Wiley-Interscience.

Monoclonal Antibodies and Fluorescent Molecules

Immunology Today. 2000. Volume 21, issue 8.

Tsien, R. Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem.* 67:509–544.

Kendall, J. M., and Badminton, M. N. 1998. *Aequorea victoria* bioluminescence moves into an exciting era. *Trends Biotechnol.* 16:216–234.

Goding, J. W. 1996. *Monoclonal Antibodies: Principles and Practice.* Academic Press.

Köhler, G., and Milstein, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256:495–497.

Chemical Synthesis of Proteins

Bang, D., Chopra, N., and Kent, S. B. 2004. Total chemical synthesis of crambin. *J. Am. Chem. Soc.* 126:1377–1383.

Dawson, P. E., and Kent, S. B. 2000. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 69:923–960.

Mayo, K. H. 2000. Recent advances in the design and construction of synthetic peptides: For the love of basics or just for the technology of it. *Trends Biotechnol.* 18:212–217.

Chapter 4

Where to Start

Felsenfeld, G. 1985. DNA. *Sci. Am.* 253(4):58–67.

Darnell, J. E., Jr. 1985. RNA. *Sci. Am.* 253(4):68–78.

Dickerson, R. E. 1983. The DNA helix and how it is read. *Sci. Am.* 249(6):94–111.

Crick, F. H. C. 1954. The structure of the hereditary material. *Sci. Am.* 191(4):54–61.

Chambon, P. 1981. Split genes. *Sci. Am.* 244(5):60–71.

Watson, J. D., and Crick, F. H. C. 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171:737–738.

Watson, J. D., and Crick, F. H. C. 1953. Genetic implications of the structure of deoxyribonucleic acid. *Nature* 171:964–967.

Meselson, M., and Stahl, F. W. 1958. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 44:671–682.

Books

Bloomfield, V. A., Crothers, D. M., Tinoco, I., and Hearst, J. 2000. *Nucleic Acids: Structures, Properties, and Functions.* University Science Books.

Singer, M., and Berg, P. 1991. *Genes and Genomes: A Changing Perspective*. University Science Books.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M. Bretscher, A., Ploegh, H., Amon, A., and Scott, M. P. 2012. *Molecular Cell Biology* (7th ed.). W. H. Freeman and Company.

Krebs, J. E, Goldstein, E. S. and Kilpatrick, S. T. 2012. *Lewin's Genes XI*. (11th ed.). Jones and Bartlett.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. 2013. *Molecular Biology of the Gene* (7th ed.). Benjamin Cummings.

DNA Structure

Neidle, S. 2007. *Principles of Nucleic Acid Structure*. Academic Press.

Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. L. 1982. The anatomy of A-, B-, and Z-DNA. *Science* 216:475–485.

Sinden, R. R. 1994. *DNA Structure and Function*. Academic Press.

DNA Replication

Lehman, I. R. 2003. Discovery of DNA polymerase. *J. Biol. Chem.* 278:34733–34738.

Hübscher, U., Maga, G., and Spardari, S. 2002. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71:133–163.

Hübscher, U., Nasheuer, H.-P., and Syväoja, J. E. 2000. Eukaryotic DNA polymerases: A growing family. *Trends Biochem. Sci.* 25:143–147.

Brautigam, C. A., and Steitz, T. A. 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* 8:54–63.

Kornberg, A. 2005. *DNA Replication* (2d ed.). University Science Books.

Discovery of Messenger RNA

Jacob, F., and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318–356.

Brenner, S., Jacob, F., and Meselson, M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576–581.

Hall, B. D., and Spiegelman, S. 1961. Sequence complementarity of T2-DNA and T2-specific RNA. *Proc. Natl. Acad. Sci. U.S.A.* 47:137–146.

Genetic Code

Novoa, E. M., and Ribas de Pouplana, L. 2012. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet.* 28:574–581.

Koonin, E. V., and Novozhilov, A. S. 2009. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* 61:99–111.

Yarus, M., Caporaso, J. G., and Knight, R. 2005. Origins of the genetic code: The escaped triplet theory. *Annu. Rev. Biochem.* 74:179–198.

Freeland, S. J., and Hurst, L. D. 2004. Evolution encoded. *Sci. Am.* 290(4):84–91.

Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. 1961. General nature of the genetic code for proteins. *Nature* 192:1227–1232.

Knight, R. D., Freeland, S. J., and Landweber L. F. 1999. Selection, history and chemistry: The three faces of the genetic code. *Trends Biochem. Sci.* 24(6):241–247.

Introns, Exons, and Split Genes

Liu, M., and Grigoriev, A. 2004. Protein domains correlate strongly with exons in multiple eukaryotic genomes—evidence of exon shuffling? *Trends Genet.* 20:399–403.

Dorit, R. L., Schoenbach, L., and Gilbert, W. 1990. How big is the universe of exons? *Science* 250:1377–1382.

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P. 1979. Organization and sequence studies of the 17-piece chicken conalbumin gene. *Nature* 282:567–574.

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., and Leder, P. 1978. Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. *Proc. Natl. Acad. Sci. U.S.A.* 75:725–729.

Reminiscences and Historical Accounts

Watson, J. D., Gann, A., and Witkowski, J. (Eds.). 2012. *The Annotated and Illustrated Double Helix*. Simon and Shuster.

Nirenberg, M. 2004. Deciphering the genetic code—a personal account. *Trends Biochem. Sci.* 29:46–54.

Clayton, J., and Dennis, C. (Eds.). 2003. *50 Years of DNA*. Palgrave Macmillan.

Watson, J. D. 1968. *The Double Helix*. Atheneum.

McCarty, M. 1985. *The Transforming Principle: Discovering That Genes Are Made of DNA*. Norton.

Cairns, J., Stent, G. S., and Watson, J. D. 2000. *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory.

Olby, R. 1974. *The Path to the Double Helix*. University of Washington Press.

Judson, H. F. 1996. *The Eighth Day of Creation*. Cold Spring Harbor Laboratory.

Sayre, A. 2000. *Rosalind Franklin and DNA*. Norton.

Chapter 5

Where to Start

Berg, P. 1981. Dissections and reconstructions of genes and chromosomes. *Science* 213:296–303.

Gilbert, W. 1981. DNA sequencing and gene structure. *Science* 214:1305–1312.

Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science* 214:1205–1210.

Mullis, K. B. 1990. The unusual origin of the polymerase chain reaction. *Sci. Am.* 262(4):56–65.

Books on Recombinant DNA Technology

Watson, J. D., Myers, R. M., Caudy, A. A., and Witkowski, J. 2007. *Recombinant DNA: Genes and Genomes* (3d ed.). W. H. Freeman and Company.

Grierson, D. (Ed.). 1991. *Plant Genetic Engineering*. Chapman and Hall.

Mullis, K. B., Ferré, F., and Gibbs, R. A. (Eds.). 1994. *The Polymerase Chain Reaction*. Birkhaüser.

Green, M. R., and Sambrook, S. 2014. *Molecular Cloning: A Laboratory Manual* (4th ed.). Cold Spring Harbor Laboratory Press.

Ausubel, F. M., Brent, R., Kingston, R. E., and Moore, D. D. (Eds.). 2002. *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*. Wiley.

Birren, B., Green, E. D., Klapholz, S., Myers, R. M., Roskams, J., Riethamn, H., and Hieter, P. (Eds.). 1999. *Genome Analysis* (vols. 1–4). Cold Spring Harbor Laboratory Press.

Methods in Enzymology. Academic Press. [Many volumes in this series deal with recombinant DNA technology.]

DNA Sequencing and Synthesis

Hunkapiller, T., Kaiser, R. J., Koop, B. F., and Hood, L. 1991. Large-scale and automated DNA sequence determination. *Science* 254:59–67.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74:5463–5467.

Maxam, A. M., and Gilbert, W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:560–564.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–679.

Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holmes, C. P., and Fodor, S. P. A. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. U.S.A.* 91:5022–5026.

Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. 1998. Shotgun sequencing of the human genome. *Science* 280:1540–1542.

Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387–402.

Metzker, M. L. 2010. Sequencing technologies—the next generation. *Nature Rev. Genet.* 11:31–46.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., et al. 2011. An integrated semi-conductor device enabling non-optical genome sequencing. *Nature*. 475:348–352.

Polymerase Chain Reaction

Arnheim, N., and Erlich, H. 1992. Polymerase chain reaction strategy. *Annu. Rev. Biochem.* 61:131–156.

Kirby, L. T. (Ed.). 1997. *DNA Fingerprinting: An Introduction.* Stockton Press.

Eisenstein, B. I. 1990. The polymerase chain reaction: A new method for using molecular genetics for medical diagnosis. *New Engl. J. Med.* 322:178–183.

Foley, K. P., Leonard, M. W., and Engel, J. D. 1993. Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* 9:380–386.

Pääbo, S. 1993. Ancient DNA. *Sci. Am.* 269(5):86–92.

Hagelberg, E., Gray, I. C., and Jeffreys, A. J. 1991. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* 352:427–429.

Lawlor, D. A., Dickel, C. D., Hauswirth, W. W., and Parham, P. 1991. Ancient HLA genes from 7500-year-old archaeological remains. *Nature* 349:785–788.

Krings, M., Geisert, H., Schmitz, R. W., Krainitzki, H., and Pääbo, S. 1999. DNA sequence of the mitochondrial hypervariable region II for the Neanderthal type specimen. *Proc. Natl. Acad. Sci. U.S.A.* 96:5581–5585.

Ovchinnikov, I. V., Götherström, A., Romanova, G. P., Kharitonov, V. M., Lidén, K., and Goodwin, W. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404: 490–493.

Genome Sequencing

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. 2001. The sequence of the human genome. *Science* 291:1304–1351.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.

Koonin, E. V. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1:127–236.

Gilligan, P., Brenner, S., and Venkatesh, B. 2002. Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences. *Gene* 294:35–44.

Enard, W., and Pääbo, S. 2004. Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* 5:351–378.

Quantitative PCR and DNA Arrays

Duggan, D. J., Bittner, J. M., Chen, Y., Meltzer, P., and Trent, J. M. 1999. Expression profiling using cDNA microarrays. *Nat. Genet.* 21:10–14.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., et al. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffery, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. 2000. Molecular portraits of human breast tumours. *Nature* 406:747–752.

Walker, N. J. 2002. A technique whose time has come. *Science* 296:557–559.

Manipulation of Eukaryotic Genes

Anderson, W. F. 1992. Human gene therapy. *Science* 256:808–813.

Friedmann, T. 1997. Overcoming the obstacles to gene therapy. *Sci. Am.* 277(6):96–101.

Blaese, R. M. 1997. Gene therapy for cancer. *Sci. Am.* 277(6):111–115.

Brinster, R. L., and Palmiter, R. D. 1986. Introduction of genes into the germ lines of animals. *Harvey Lect.* 80:1–38.

Capecchi, M. R. 1989. Altering the genome by homologous recombination. *Science* 244:1288–1292.

Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N., and Klein, W. H. 1993. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* 364:501–506.

Parkmann, R., Weinberg, K., Crooks, G., Nolta, J., Kapoor, N., and Kohn, D. 2000. Gene therapy for adenosine deaminase deficiency. *Annu. Rev. Med.* 51:33–47.

Gaj, T., Gersbach, C. A., and Barbas III, C. F. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31:397–405.

RNA Interference

Rana, T. M. 2007. Illuminating the silence: Understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* 8:23–36.

Novina, C. D., and Sharp, P. A. 2004. The RNAi revolution. *Nature* 430:161–164.

Hannon, G. J., and Rossi, J. J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* 431:371–378.

Meister, G., and Tuschl, T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* 431:343–349.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411:494–498.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.

Genetic Engineering of Plants

Gasser, C. S., and Fraley, R. T. 1992. Transgenic crops. *Sci. Am.* 266(6):62–69.

Gasser, C. S., and Fraley, R. T. 1989. Genetically engineering plants for crop improvement. *Science* 244:1293–1299.

Shimamoto, K., Terada, R., Izawa, T., and Fujimoto, H. 1989. Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* 338:274–276.

Chilton, M.-D. 1983. A vector for introducing new genes into plants. *Sci. Am.* 248(6):50–59.

Hansen, G., and Wright, M. S. 1999. Recent advances in the transformation of plants. *Trends Plant Sci.* 4:226–231.

Hammond, J. 1999. Overview: The many uses of transgenic plants. *Curr. Top. Microbiol. Immunol.* 240:1–20.

Finer, J. J., Finer, K. R., and Ponappa, T. 1999. Particle bombardment mediated transformation. *Curr. Top. Microbiol. Immunol.* 240: 60–80.

Amyotrophic Lateral Sclerosis

Siddique, T., Figlewicz, D. A., Pericak-Vance, M. A., Haines, J. L., Rouleau, G., Jeffers, A. J., Sapp, P., Hung, W.-Y., Bebout, J., McKenna-Yasek, D., et al. 1991. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. *New Engl. J. Med.* 324:1381–1384.

Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J. P., Deng, H.-X., et al. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362:59–62.

Gurney, M. E., Pu, H., Chiu, A. Y., Dal Canto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H.-X., et al. 1994. Motor neuron degeneration in mice that express a human Cu, Zn superoxide dismutase mutation. *Science* 264:1772–1774.

Borchelt, D. R., Lee, M. K., Slunt, H. S., Guarnieri, M., Xu, Z.-S., Wong, P. C., Brown, R. H., Jr., Price, D. L., Sisodia, S. S., and Cleveland, D. W. 1994. Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc. Natl. Acad. Sci. U.S.A.* 91:8292–8296.

Chapter 6

Where to Start

Hogewer, P. 2011. The roots of bioinformatics in theoretical biology. *PLoS Comp. Biol.* 7:e1002021.

Searls, D. B. 2010. The roots of bioinformatics. *PLoS Comp. Biol.* 6: e1000809.

Books

Claverie, J.-M., and Notredame, C. 2003. *Bioinformatics for Dummies*. Wiley.

Pevsner, J. 2003. *Bioinformatics and Functional Genomics*. Wiley-Liss.

Doolittle, R. F. 1987. *Of URFS and ORFS*. University Science Books.

Sequence Alignment

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29: 2994–3005.

Henikoff, S., and Henikoff, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89:10915–10919.

Johnson, M. S., and Overington, J. P. 1993. A structural basis for sequence comparisons: An evaluation of scoring methodologies. *J. Mol. Biol.* 233:716–738.

Eddy, S. R. 2004. Where did the BLOSUM62 alignment score matrix come from? *Nat. Biotechnol.* 22:1035–1036.

Aravind, L., and Koonin, E. V. 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* 287:1023–1040.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Structure Comparison

Orengo, C. A., Bray, J. E., Buchan, D. W., Harrison, A., Lee, D., Pearl, F. M., Sillitoe, I., Todd, A. E., and Thornton, J. M. 2002. The CATH protein family database: A resource for structural and functional annotation of genomes. *Proteomics* 2:11–21.

Bashford, D., Chothia, C., and Lesk, A. M. 1987. Determinants of a protein fold: Unique features of the globin amino acid sequences. *J. Mol. Biol.* 196:199–216.

Harutyunyan, E. H., Safonova, T. N., Kuranova, I. P., Popov, A. N., Teplyakov, A. V., Obmolova, G. V., Rusakov, A. A., Vainshtein, B. K., Dodson, G. G., Wilson, J. C., et al. 1995. The structure of deoxy- and oxy-leghaemoglobin from lupin. *J. Mol. Biol.* 251:104–115.

Flaherty, K. M., McKay, D. B., Kabsch, W., and Holmes, K. C. 1991. Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc. Natl. Acad. Sci. U.S.A.* 88:5041–5045.

Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

Hadley, C., and Jones, D. T. 1999. A systematic comparison of protein structure classification: SCOP, CATH and FSSP. *Struct. Fold. Des.* 7:1099–1112.

Domain Detection

Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., et al. 2003. CDD: A curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* 31:383–387.

Ploegman, J. H., Drent, G., Kalk, K. H., and Hol, W. G. 1978. Structure of bovine liver rhodanese I: Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. *J. Mol. Biol.* 123:557–594.

Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., and Burley, S. K. 1992. Crystal structure of TFIID TATA-box binding protein. *Nature* 360:40–46.

Doolittle, R. F. 1995. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* 64:287–314.

Heger, A., and Holm, L. 2000. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* 41:224–237.

Evolutionary Trees

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. 2002. Genome trees and the tree of life. *Trends Genet.* 18:472–479.

Doolittle, R. F. 1992. Stein and Moore Award address. Reconstructing history with amino acid sequences. *Protein Sci.* 1:191–200.

Zuckerkandl, E., and Pauling, L. 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.* 8:357–366.

Schönknecht, G., Chen, W.-H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Bräutigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., et al. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339:1207–1210.

Ancient DNA

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo C., et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226.

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., et al. 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134:416–426.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. 2004. Genetic analyses from ancient DNA. *Annu. Rev. Genet.* 38:645–679.

Evolution in the Laboratory

Sassanfar, M., and Szostak, J. W. 1993. An RNA motif that binds ATP. *Nature* 364:550–553.

Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. 1995. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64:763–797.

Wilson, D. S., and Szostak, J. W. 1999. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* 68:611–647.

Hermann, T., and Patel, D. J. 2000. Adaptive recognition by nucleic acid aptamers. *Science* 287:820–825.

Keefe, A. D., Pai, S., and Ellington, A. 2010. Aptamers as therapeutics. *Nat. Rev. Drug Discov.* 9:537–550.

Radom, F., Jurek, P. M., Mazurek, M. P., Otlewski, J., and Jeleń, F. 2013. Aptamers: Molecules of great potential. *Biotechnol. Adv.* 31:1260–1274.

Web Sites

The Protein Data Bank (PDB) site is the repository for three-dimensional macromolecular structures. It currently contains more than 100,000 structures. (http://www.pdb.org).

National Center for Biotechnology Information (NCBI) contains molecular biological databases and software for analysis. (http://www.ncbi.nlm.nih.gov/).

Chapter 7

Where to Start

Changeux, J.-P. 2011. 50th anniversary of the word "Allosteric." *Protein Sci.* 20:1119–1124.

Perutz, M. F. 1978. Hemoglobin structure and respiratory transport. *Sci. Am.* 239(6):92–125.

Perutz, M. F. 1980. Stereochemical mechanism of oxygen transport by haemoglobin. *Proc. R. Soc. Lond. Biol. Sci.* 208:135–162.

Kilmartin, J. V. 1976. Interaction of haemoglobin with protons, CO_2, and 2,3-diphosphoglycerate. *Brit. Med. Bull.* 32:209–222.

Structure

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181:662–666.

Shaanan, B. 1983. Structure of human oxyhaemoglobin at 2.1 Å resolution. *J. Mol. Biol.* 171:31–59.

Frier, J. A., and Perutz, M. F. 1977. Structure of human foetal deoxyhaemoglobin. *J. Mol. Biol.* 112:97–112.

Perutz, M. F. 1969. Structure and function of hemoglobin. *Harvey Lect.* 63:213–261.

Perutz, M. F. 1962. Relation between structure and sequence of haemoglobin. *Nature* 194:914–917.

Harrington, D. J., Adachi, K., and Royer, W. E., Jr. 1997. The high resolution crystal structure of deoxyhemoglobin S. *J. Mol. Biol.* 272:398–407.

Interaction of Hemoglobin with Allosteric Effectors

Benesch, R., and Beesch, R. E. 1969. Intracellular organic phosphates as regulators of oxygen release by haemoglobin. *Nature* 221: 618–622.

Fang, T. Y., Zou, M., Simplaceanu, V., Ho, N. T., and Ho, C. 1999. Assessment of roles of surface histidyl residues in the molecular basis of the Bohr effect and of β 143 histidine in the binding of 2,3-bisphosphoglycerate in human normal adult hemoglobin. *Biochemistry* 38:13423–13432.

Arnone, A. 1992. X-ray diffraction study of binding of 2,3-diphosphoglycerate to human deoxyhaemoglobin. *Nature* 237:146–149.

Models for Cooperativity

Changeux, J.-P. 2012. Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu. Rev. Biophys.* 41:103–133.

Monod, J., Wyman, J., and Changeux, J.-P. 1965. On the nature of allosteric interactions: A plausible model. *J. Mol. Biol.* 12: 88–118.

Koshland, D. L., Jr., Nemethy, G., and Filmer, D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5:365–385.

Ackers, G. K., Doyle, M. L., Myers, D., and Daugherty, M. A. 1992. Molecular code for cooperativity in hemoglobin. *Science* 255: 54–63.

Sickle-Cell Anemia and Thalassemia

Herrick, J. B. 1910. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. *Arch. Intern. Med.* 6: 517–521.

Pauling, L., Itano, H. A., Singer, S. J., and Wells, L. C. 1949. Sickle cell anemia: A molecular disease. *Science* 110:543–548.

Ingram, V. M. 1957. Gene mutation in human hemoglobin: The chemical difference between normal and sickle cell haemoglobin. *Nature* 180:326–328.

Eaton, W. A., and Hofrichter, J. 1990. Sickle cell hemoglobin polymerization. *Adv. Prot. Chem.* 40:63–279.

Weatherall, D. J. 2001. Phenotype genotype relationships in monogenic disease: Lessons from the thalassemias. *Nat. Rev. Genet.* 2:245–255.

Tsaras, G., Owusu-Ansah, A., Boateng, F. O., and Amoateng-Adjepong, Y. 2009. Complications associated with sickle cell trait: A brief narrative review. *Am. J. Med.* 122:507–512.

Globin-Binding Proteins and Other Globins

Helbo, S., Weber, R. E., and Fago, A. 2013. Expression patterns and adaptive functional diversity of vertebrate myoglobins. *Biochim. Biophys. Acta* 1834:1832–1839.

Kihm, A. J., Kong, Y., Hong, W., Russell, J. E., Rouda, S., Adachi, K., Simon, M. C., Blobel, G. A., and Weiss, M. J. 2002. An abundant erythroid protein that stabilizes free α-haemoglobin. *Nature* 417:758–763.

Feng, L., Zhou, S., Gu, L., Gell, D. A., Mackay, J. P., Weiss, M. J., Gow, A. J., and Shi, Y. 2005. Structure of oxidized α-haemoglobin bound to AHSP reveals a protective mechanism for haem. *Nature* 435:697–701.

Yu, X., Kong, Y., Dore, L. C., Abdulmalik, O., Katein, A. M., Zhou, S., Choi, J. K., Gell, D., Mackay, J. P., Gow, A. J., et al. 2007. An erythroid chaperone that facilitates folding of α-globin subunits for hemoglobin synthesis. *J. Clin. Invest.* 117:1856–1865.

Burmester, T., Haberkamp, M., Mitz, S., Roesner, A., Schmidt, M., Ebner, B., Gerlach, F., Fuchs, C., and Hankeln, T. 2004. Neuroglobin and cytoglobin: Genes, proteins and evolution. *IUBMB Life* 56:703–707.

Hankeln, T., Ebner, B., Fuchs, C., Gerlach, F., Haberkamp, M., Laufs, T. L., Roesner, A., Schmidt, M., Weich, B., Wystub, S., et al. 2005. Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J. Inorg. Biochem.* 99:110–119.

Burmester, T., Ebner, B., Weich, B., and Hankeln, T. 2002. Cytoglobin: A novel globin type ubiquitously expressed in vertebrate tissues. *Mol. Biol. Evol.* 19:416–421.

Zhang, C., Wang, C., Deng, M., Li, L., Wang, H., Fan, M., Xu, W., Meng, F., Qian, L., and He, F. 2002. Full-length cDNA cloning of human neuroglobin and tissue expression of rat neuroglobin. *Biochem. Biophys. Res. Commun.* 290:1411–1419.

Chapter 8

Where to Start

Zalatan, J. G., and Herschlag, D. 2009. The far reaches of enzymology. *Nat. Chem. Biol.* 5:516–520.

Hammes, G. G. 2008. How do enzymes really work? *J. Biol. Chem.* 283:22337–22346.

Koshland, D. E., Jr. 1987. Evolution of catalytic function. *Cold Spring Harbor Symp. Quant. Biol.* 52:1–7.

Jencks, W. P. 1987. Economics of enzyme catalysis. *Cold Spring Harbor Symp. Quant. Biol.* 52:65–73.

Lerner, R. A., and Tramontano, A. 1988. Catalytic antibodies. *Sci. Am.* 258(3):58–70.

Books

Cook, P. F., and Cleland, W. W. 2007. *Enzyme Kinetics and Mechanism.* Garland Press.

Fersht, A. 1999. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* W. H. Freeman and Company.

Walsh, C. 1979. *Enzymatic Reaction Mechanisms.* W. H. Freeman and Company.

Bender, M. L., Bergeron, R. J., and Komiyama, M. 1984. *The Bioorganic Chemistry of Enzymatic Catalysis.* Wiley-Interscience.

Abelson, J. N., and Simon, M. I. (Eds.). 1992. *Methods in Enzymology.* Academic Press.

Friedmann, H. C. (Ed.). 1981. *Benchmark Papers in Biochemistry,* vol. 1, *Enzymes.* Hutchinson Ross.

Transition-State Stabilization, Analogs, and Other Enzyme Inhibitors

Schramm, V. L. 2007. Enzymatic transition state theory and transition state analog design. *J. Biol. Chem.* 282:28297–28300.

Pauling, L. 1948. Nature of forces between large molecules of biological interest. *Nature* 161:707–709.

Leinhard, G. E. 1973. Enzymatic catalysis and transition-state theory. *Science* 180:149–154.

Kraut, J. 1988. How do enzymes work? *Science* 242:533–540.

Waxman, D. J., and Strominger, J. L. 1983. Penicillin-binding proteins and the mechanism of action of β-lactam antibiotics. *Annu. Rev. Biochem.* 52:825–869.

Abraham, E. P. 1981. The β-lactam antibiotics. *Sci. Am.* 244(6):76–86.

Walsh, C. T. 1984. Suicide substrates, mechanism-based enzyme inactivators: Recent developments. *Annu. Rev. Biochem.* 53:493–535.

Catalytic Antibodies

Hilvert, D. 2000. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* 69:751–794.

Wade, H., and Scanlan, T. S. 1997. The structural and functional basis of antibody catalysis. *Annu. Rev. Biophys. Biomol. Struct.* 26:461–493.

Lerner, R. A., Benkovic, S. J., and Schultz, P. G. 1991. At the crossroads of chemistry and immunology: Catalytic antibodies. *Science* 252:659–667.

Cochran, A. G., and Schultz, P. G. 1990. Antibody-catalyzed porphyrin metallation. *Science* 249:781–783.

Enzyme Kinetics and Mechanisms

Hammes, G. G., Benkovic, S. J., and Hammes-Schiffer, S. 2011. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochemistry* 50:10422–10430.

Johnson, K. A., and Goody, R. S. 2011. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* 50:8264–8269.

Hammes-Schiller, S., and Benkovic, S. J. 2006. Relating protein motion to catalysis. *Annu. Rev. Biochem.* 75:519–541.

Benkovic, S. J., and Hammes-Schiller, S. 2003. A perspective on enzyme catalysis. *Science* 301:1196–1202.

Hur, S., and Bruice, T. C. 2003. The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc. Natl. Acad. Sci. U.S.A.* 100:12015–12020.

Miles, E. W., Rhee, S., and Davies, D. R. 1999. The molecular basis of substrate channeling. *J. Biol. Chem.* 274:12193–12196.

Warshel, A. 1998. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* 273:27035–27038.

Cannon, W. R., and Benkovic, S. J. 1999. Solvation, reorganization energy, and biological catalysis. *J. Biol. Chem.* 273:26257–26260.

Cleland, W. W., Frey, P. A., and Gerlt, J. A. 1998. The low barrier hydrogen bond in enzymatic catalysis. *J. Biol. Chem.* 273: 25529–25532.

Romesberg, F. E., Santarsiero, B. D., Spiller, B., Yin, J., Barnes, D., Schultz, P. G., and Stevens, R. C. 1998. Structural and kinetic evidence for strain in biological catalysis. *Biochemistry* 37:14404–14409.

Fersht, A. R., Leatherbarrow, R. J., and Wells, T. N. C. 1986. Binding energy and catalysis: A lesson from protein engineering of the tyrosyl-tRNA synthetase. *Trends Biochem. Sci.* 11:321–325.

Jencks, W. P. 1975. Binding energy, specificity, and enzymic catalysis: The Circe effect. *Adv. Enzymol.* 43:219–410.

Knowles, J. R., and Albery, W. J. 1976. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* 15:5631–5640.

Single Molecule Studies

Allewell, N. M. 2010. Thematic Minireview Series: Single-molecule Measurements in Biochemistry and Molecular Biology. *J. Biol. Chem.* 285:18959–18983. A series of reviews on single-molecule studies.

Min, W., English, B. P., Lou, G., Cherayil, B. J., Kou, S. C., and Xie, X. S. 2005. Fluctuating Enzymes: Lessons from Single-Molecule Studies. *Acc. Chem. Res.* 38: 923–931.

Xie, X. S., and Lu, H. P. 1999. Single-molecule enzymology. *J. Biol. Chem.* 274:15967–15970.

Lu, H. P., Xun, L., and Xie, X. S. 1998. Single-molecule enzymatic dynamics. *Science* 282:1877–1882.

Chapter 9

Where to Start

Stroud, R. M. 1974. A family of protein-cutting proteins. *Sci. Am.* 231(1):74–88.

Kraut, J. 1977. Serine proteases: Structure and mechanism of catalysis. *Annu. Rev. Biochem.* 46:331–358.

Lindskog, S. 1997. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* 74:1–20.

Jeltsch, A., Alves, J., Maass, G., and Pingoud, A. 1992. On the catalytic mechanism of EcoRI and EcoRV: A detailed proposal based on biochemical results, structural data and molecular modelling. *FEBS Lett.* 304:4–8.

Bauer, C. B., Holden, H. M., Thoden, J. B., Smith, R., and Rayment, I. 2000. X-ray structures of the apo and MgATP-bound states of *Dictyostelium discoideum* myosin motor domain. *J. Biol. Chem.* 275:38494–38499.

Lolis, E., and Petsko, G. A. 1990. Transition-state analogues in protein crystallography: Probes of the structural source of enzyme catalysis. *Annu. Rev. Biochem.* 59:597–630.

Books

Fersht, A. 1999. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* W. H. Freeman and Company.

Silverman, R. B. 2000. *The Organic Chemistry of Enzyme-Catalyzed Reactions*. Academic Press.

Page, M., and Williams, A. 1997. *Organic and Bio-organic Mechanisms*. Addison Wesley Longman.

Chymotrypsin and Other Serine Proteases

Fastrez, J., and Fersht, A. R. 1973. Demonstration of the acyl-enzyme mechanism for the hydrolysis of peptides and anilides by chymotrypsin. *Biochemistry* 12:2025–2034.

Sigler, P. B., Blow, D. M., Matthews, B. W., and Henderson, R. 1968. Structure of crystalline-chymotrypsin II: A preliminary report including a hypothesis for the activation mechanism. *J. Mol. Biol.* 35:143–164.

Kossiakoff, A. A., and Spencer, S. A. 1981. Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: Neutron structure of trypsin. *Biochemistry* 20:6462–6474.

Carter, P., and Wells, J. A. 1988. Dissecting the catalytic triad of a serine protease. *Nature* 332:564–568.

Carter, P., and Wells, J. A. 1990. Functional interaction among catalytic residues in subtilisin BPN′. *Proteins* 7:335–342.

Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. 2000. Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 Å resolution: Structural basis of Janus-faced serine protease inhibitor specificity. *J. Mol. Biol.* 298:477–491.

Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G. J., and Fontecilla-Camps, J. C. 2000. Crystal structure of the catalytic domain of human complement C1s: A serine protease with a handle. *EMBO J.* 19:1755–1765.

Bachovchin D. A., and Cravatt B. F. 2012. The pharmacological landscape and therapeutic potential of serine hydrolases. *Nature Reviews Drug Discovery*. 11:52–68.

Other Proteases

Vega, S., Kang, L. W., Velazquez-Campoy, A., Kiso, Y., Amzel, L. M., and Freire, E. 2004. A structural and thermodynamic escape mechanism from a drug resistant mutation of the HIV-1 protease. *Proteins* 55:594–602.

Kamphuis, I. G., Kalk, K. H., Swarte, M. B., and Drenth, J. 1984. Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* 179:233–256.

Kamphuis, I. G., Drenth, J., and Baker, E. N. 1985. Thiol proteases: Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* 182:317–329.

Sivaraman, J., Nagler, D. K., Zhang, R., Menard, R., and Cygler, M. 2000. Crystal structure of human procathepsin X: A cysteine protease with the proregion covalently linked to the active site cysteine. *J. Mol. Biol.* 295:939–951.

Davies, D. R. 1990. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* 19:189–215.

Dorsey, B. D., Levin, R. B., McDaniel, S. L., Vacca, J. P., Guare, J. P., Darke, P. L., Zugay, J. A., Emini, E. A., Schleif, W. A., Quintero, J. C., et al. 1994. L-735,524: The design of a potent and orally bio-available HIV protease inhibitor. *J. Med. Chem.* 37:3443–3451.

Chen, Z., Li, Y., Chen, E., Hall, D. L., Darke, P. L., Culberson, C., Shafer, J. A., and Kuo, L. C. 1994. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* 269:26344–26348.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., et al. 1992. The α/β hydrolase fold. *Protein Eng.* 5:197–211.

Carbonic Anhydrase

Lindskog, S., and Coleman, J. E. 1973. The catalytic mechanism of carbonic anhydrase. *Proc. Natl. Acad. Sci. U.S.A.* 70:2505–2508.

Kannan, K. K., Notstrand, B., Fridborg, K., Lovgren, S., Ohlsson, A., and Petef, M. 1975. Crystal structure of human erythrocyte carbonic anhydrase B: Three-dimensional structure at a nominal 2.2-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 72:51–55.

Boriack-Sjodin, P. A., Zeitlin, S., Chen, H. H., Crenshaw, L., Gross, S., Dantanarayana, A., Delgado, P., May, J. A., Dean, T., and Christianson, D. W. 1998. Structural analysis of inhibitor binding to human carbonic anhydrase II. *Protein Sci.* 7:2483–2489.

Wooley, P. 1975. Models for metal ion function in carbonic anhydrase. *Nature* 258:677–682.

Jonsson, B. H., Steiner, H., and Lindskog, S. 1976. Participation of buffer in the catalytic mechanism of carbonic anhydrase. *FEBS Lett.* 64:310–314.

Sly, W. S., and Hu, P. Y. 1995. Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* 64:375–401.

Maren, T. H. 1988. The kinetics of HCO_3^- synthesis related to fluid secretion, pH control, and CO_2 elimination. *Annu. Rev. Physiol.* 50:695–717.

Roy, A., and Taraphder, S. 2010. Role of protein motions on proton transfer pathways in human carbonic anhydrase II. *Biochim. Biophys. Acta* 1804:352–361.

Restriction Enzymes

Selvaraj, S., Kono, H., and Sarai, A. 2002. Specificity of protein-DNA recognition revealed by structure-based potentials: Symmetric/asymmetric and cognate/non-cognate binding. *J. Mol. Biol.* 322:907–915.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K., and Wilson, K. S. 1993. The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* 12:1781–1795.

Kostrewa, D., and Winkler, F. K. 1995. Mg^{2+} binding to the active site of EcoRV endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* 34:683–696.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S., and Kokkinidis, M. 1994. Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.* 1:469–475.

Sam, M. D., and Perona, J. J. 1999. Catalytic roles of divalent metal ions in phosphoryl transfer by EcoRV endonuclease. *Biochemistry* 38:6576–6586.

Jeltsch, A., and Pingoud, A. 1996. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.* 42:91–96.

Advani S., Mishra P., Dubey S., and Thakur S. 2010. Categoric prediction of metal ion mechanisms in the active sites of 17 select type II restriction endonucleases. *Biochem. Biophys. Res. Commun.* 402:177–179.

Myosins

Grigorenko, B. L., Rogov, A. V., Topol, I. A., Burt, S. K., Martinez, H. M., and Nemukhin, A. V. 2007. Mechanism of the myosin catalyzed hydrolysis of ATP as rationalized by molecular modeling. *Proc. Natl. Acad. Sci. U.S.A.* 104:7057–7061.

Gulick, A. M., Bauer, C. B., Thoden, J. B., and Rayment, I. 1997. X-ray structures of the MgADP, MgATPγ S, and MgAMPPNP complexes of the *Dictyostelium discoideum* myosin motor domain. *Biochemistry* 36:11619–11628.

Kovacs, M., Malnasi-Csizmadia, A., Woolley, R. J., and Bagshaw, C. R. 2002. Analysis of nucleotide binding to *Dictyostelium* myosin II motor domains containing a single tryptophan near the active site. *J. Biol. Chem.* 277:28459–28467.

Kuhlman, P. A., and Bagshaw, C. R. 1998. ATPase kinetics of the *Dictyostelium discoideum* myosin II motor domain. *J. Muscle Res. Cell Motil.* 19:491–504.

Smith, C. A., and Rayment, I. 1996. X-ray structure of the magnesium(II) ADP vanadate complex of the *Dictyostelium discoideum* myosin motor domain to 1.9 Å resolution. *Biochemistry* 35:5404–5417.

Yildiz A., Forkey J. N., McKinney S. A., Ha T., Goldman Y. E., and Selvin P. R. 2003. Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization. *Science*. 300:2061–2065.

Chapter 10

Where to Start

Kyriakis, J. M. 2014. In the beginning, there was protein phosphorylation. *J. Biol. Chem.* 289:9460—9462.

Changeux, J.-P. 2011. 50th anniversary of the word “Allosteric.” *Protein Sci.* 20:1119–1124.

Kantrowitz, E. R., and Lipscomb, W. N. 1990. *Escherichia coli* aspartate transcarbamoylase: The molecular basis for a concerted allosteric transition. *Trends Biochem. Sci.* 15:53–59.

Schachman, H. K. 1988. Can a simple model account for the allosteric transition of aspartate transcarbamoylase? *J. Biol. Chem.* 263:18583–18586.

Neurath, H. 1989. Proteolytic processing and physiological regulation. *Trends Biochem. Sci.* 14:268–271.

Bode, W., and Huber, R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* 204: 433–451.

Aspartate Transcarbamoylase and Allosteric Interactions

Changeux, J.-P. 2012. Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annu. Rev. Biophys.* 41:103–133.

Peterson, A. W., Cockrell, G. M., and Kantrowitz, E. R. 2012. A second allosteric site in *Escherichia coli* aspartate transcarbamoylase. *Biochemistry* 51:4776−4778.

Rabinowitz, J. D., Hsiao, J. J., Gryncel, K. R., Kantrowitz, E. R., Feng, X.-J., Li, G., and Rabitz H. 2008. Dissecting enzyme regulation by multiple allosteric effectors: Nucleotide regulation of aspartate transcarbamoylase. *Biochemistry* 47:5881–5888.

West, J. M., Tsuruta, H., and Kantrowitz, E. R. 2004. A fluorescent probe-labeled *Escherichia coli* aspartate transcarbamoylase that monitors the allosteric conformation state. *J. Biol. Chem.* 279:945–951.

Endrizzi, J. A., Beernink, P. T., Alber, T., and Schachman, H. K. 2000. Binding of bisubstrate analog promotes large structural changes in the unregulated catalytic trimer of aspartate transcarbamoylase: Implications for allosteric regulation. *Proc. Natl. Acad. Sci. U.S.A.* 97:5077–5082.

Beernink, P. T., Endrizzi, J. A., Alber, T., and Schachman, H. K. 1999. Assessment of the allosteric mechanism of aspartate transcarbamoylase based on the crystalline structure of the unregulated catalytic subunit. *Proc. Natl. Acad. Sci. U.S.A.* 96:5388–5393.

Wales, M. E., Madison, L. L., Glaser, S. S., and Wild, J. R. 1999. Divergent allosteric patterns verify the regulatory paradigm for aspartate transcarbamoylase. *J. Mol. Biol.* 294:1387–1400.

Newell, J. O., Markby, D. W., and Schachman, H. K. 1989. Cooperative binding of the bisubstrate analog *N*-(phosphonacetyl)-L-aspartate to aspartate transcarbamoylase and the heterotropic effects of ATP and CTP. *J. Biol. Chem.* 264:2476–2481.

Stevens, R. C., Gouaux, J. E., and Lipscomb, W. N. 1990. Structural consequences of effector binding to the T state of aspartate carbamoyl-transferase: Crystal structures of the unligated and ATP- and CTP-complexed enzymes at 2.6-Å resolution. *Biochemistry* 29:7691–7701.

Gouaux, J. E., and Lipscomb, W. N. 1990. Crystal structures of phosphonoacetamide ligated T and phosphonoacetamide and malonate ligated R states of aspartate carbamoyltransferase at 2.8-Å resolution and neutral pH. *Biochemistry* 29:389–402.

Labedan, B., Boyen, A., Baetens, M., Charlier, D., Chen, P., Cunin, R., Durbeco, V., Glansdorff, N., Herve, G., Legrain, C., et al. 1999. The evolutionary history of carbamoyltransferases: A complex set of paralogous genes was already present in the last universal common ancestor. *J. Mol. Evol.* 49:461–473.

Covalent Modification

Endicott, J. A., Noble, M. E. M., and Johnson, L. N. 2012. The structural basis for control of eukaryotic protein kinases. *Annu. Rev. Biochem.* 81:587–613.

Tarrant, M. K., and Cole, P. A. 2009. The chemical biology of protein phosphorylation *Annu. Rev. Biochem.* 78:797–825.

Guarente, L. 2011. The logic linking protein acetylation and metabolism. *Cell Metab.* 14:151–153.

Guan, K-L., and Xiong, Y. 2011. Regulation of intermediary metabolism by protein acetylation. *Trends Biochem. Sci.* 36:108–116.

Johnson, L. N., and Barford, D. 1993. The effects of phosphorylation on the structure and function of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 22:199–232.

Barford, D., Das, A. K., and Egloff, M. P. 1998. The structure and mechanism of protein phosphatases: Insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* 27:133–164.

Protein Kinase A

Taylor, S. S., Ilouz, R., Zhang, P., and Kornev, A. P. 2012. Assembly of allosteric macromolecular switches: Lessons from PKA. *Nature Rev. Mol. Cell Biol.* 13:646–658.

Zhang, P., Smith-Nguyen, E. V., Keshwani, M. M., Deal, M. S., Kornev, A. P., and Taylor, S. S. 2012. Structure and allostery of the PKA RIIb tetrameric holoenzyme. *Science* 334:712–716.

Taylor, S. S., and Kornev, A. P. 2011. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* 36:65–77.

Pearlman, S. M., Serber, Z., and Ferrell Jr., J. E. 2011. A mechanism for the evolution of phosphorylation sites. *Cell* 147:934–946.

Knighton, D. R., Zheng, J. H., TenEyck, L., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. 1991. Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253:414–420.

Zymogen Activation

Artenstein, A. W., and Opal, S. M. 2011. Proprotein convertases in health and disease. *New Engl. J. Med.* 65:2507–2518.

Neurath, H. 1986. The versatility of proteolytic enzymes. *J. Cell. Biochem.* 32:35–49.

Bode, W., and Huber, R. 1986. Crystal structure of pancreatic serine endopeptidases. In *Molecular and Cellular Basis of Digestion* (pp. 213–234), edited by P. Desnuelle, H. Sjostrom, and O. Noren. Elsevier.

James, M. N. 1991. Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* 219:671–692.

Protease Inhibitors

Gooptu, B., and Lomas, D. A. 2009. Conformational Pathology of the Serpins: Themes, Variations, and Therapeutic Strategies. *Annu. Rev. Biochem.* 78:147–167.

Carrell, R., and Travis, J. 1985. α_1-Antitrypsin and the serpins: Variation and countervariation. *Trends Biochem. Sci.* 10:20–24.

Carp, H., Miller, F., Hoidal, J. R., and Janoff, A. 1982. Potential mechanism of emphysema: α_1-Proteinase inhibitor recovered from lungs of cigarette smokers contains oxidized methionine and has decreased elastase inhibitory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 79:2041–2045.

Owen, M. C., Brennan, S. O., Lewis, J. H., and Carrell, R. W. 1983. Mutation of antitrypsin to antithrombin. *New Engl. J. Med.* 309: 694–698.

Travis, J., and Salvesen, G. S. 1983. Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* 52:655–709.

Clotting Cascade

Kollman, J. M., Pandi, L., Sawaya, M. R., Riley, M., and Doolittle, R. F. 2009. Crystal structure of human fibrinogen. *Biochemistry* 48: 3877–3886.

Furie, B., and Furie, B. C. 2008. Mechanisms of thrombus formation. *New Engl. J. Med.* 359:938–949.

Orfeo, T., Brufatto, N., Nesheim, M. E., Xu, H., Butenas, S., and Mann, K. G. 2004. The factor V activation paradox. *J. Biol. Chem.* 279:19580–19591.

Mann, K. G. 2003. Thrombin formation. *Chest* 124:4S–10S.

Rose, T., and Di Cera, E. 2002. Three-dimensional modeling of thrombin–fibrinogen interaction. *J. Biol. Chem.* 277:18875–18880.

Krem, M. M., and Di Cera, E. 2002. Evolution of cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* 27:67–74.

Fuentes-Prior, P., Iwanaga, Y., Huber, R., Pagila, R., Rumennik, G., Seto, M., Morser, J., Light, D. R., and Bode, W. 2000. Structural basis for the anticoagulant activity of the thrombin–thrombomodulin complex. *Nature* 404:518–525.

Lawn, R. M., and Vehar, G. A. 1986. The molecular genetics of hemophilia. *Sci. Am.* 254(3):48–65.

Chapter 11

Where to Start

Glycochemistry and glycobiology. A series of review articles. 2007. *Nature* 446:999–1051.

Maeder, T. 2002. Sweet medicines. *Sci. Am.* 287(1):40–47.

Freeze, H. H. 2013. Understanding human glycosylation disorders. *J. Biol. Chem.* 288:6936–6945.

Coutinho, M. F., Prata M., J., and Alves, S. 2012. Mannose-6-phosphate pathway: A review on its role in lysosomal function and dysfunction. *Mol. Gen. Metab.* 105:542–550.

Books

Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C., Hart, G., and Etzler, M. E. 2009. *Essentials of Glycobiology* (2d ed.). Cold Spring Harbor Laboratory Press.

Stick, R. V., and Williams, S. 2008. *Carbohydrates: The Essential Molecules of Life* (2d ed.). Elsevier Science.

Sansome, C., and Markman, O. 2007. *Glycobiology*. Scion.

Lindhorst, T. K. 2007. *Essentials of Carbohydrate Chemistry and Biochemistry* (3d ed.). Wiley-VCH.

Taylor, M. E. 2006. *Introduction to Glycobiology* (2d ed.). Oxford University Press.

Glycoproteins

Tran, D. T., and Hagen, K. G. T. 2013. Mucin-type O-glycosylation during development. *J. Biol. Chem.* 288:6921–6929.

Gill, D. J., Clausen H., and Bard, F. 2011. Location, location, location: New insights into O-GalNAc protein glycosylation. *Trends Cell Biol.* 21:149–158.

Foley, R. N. 2008. Erythropoietin: Physiology and molecular mechanisms. *Heart Failure Rev.* 13:404–414.

Fisher, J. W. 2003. Erythropoietin: Physiology and pharmacology update. *Exp. Biol. Med.* 228:1–14.

Cheetham, J. C., Smith, D. M., Aoki, K. H., Stevenson, J. L., Hoeffel, T. J., Syed, R. S., Egrie, J., and Harvey, T. S. 1998. NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat. Struct. Biol.* 5:861–866.

Hattrup, C. L., and Gendler, S. J. 2008. Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* 70:431–457.

Thorton, D. J., Rousseau, K., and McGuckin, M. A. 2008. Structure and function of mucins in airways mucus. *Annu. Rev. Physiol.* 70:459–486.

Rose, M. C., and Voynow, J. A. 2007. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol. Rev.* 86:245–278.

Lamoureux, F., Baud'huin, M., Duplomb, L., Heymann, D., and Rédini, F. 2007. Proteoglycans: Key partners in bone cell biology. *Bioessays* 29:758–771.

Carraway, K. L., Funes, M., Workman, H. C., and Sweeney, C. 2007. Contribution of membrane mucins to tumor progression through modulation of cellular growth signaling pathways. *Curr. Top. Dev. Biol.* 78:1–22.

Yan, A., and Lennarz, W. J. 2005. Unraveling the mechanism of protein *N*-glycosylation. *J. Biol. Chem.* 280:3121–3124.

Pratta, M. A., Yao, W., Decicco, C., Tortorella, M., Liu, R.-Q., Copeland, R. A., Magolda, R., Newton, R. C., Trzaskos, J. M., and Arner, E. C. 2003. Aggrecan protects cartilage collagen from proteolytic cleavage. *J. Biol. Chem.* 278:45539–45545.

Glycosyltransferases

Wells, L. 2013. The O-mannosylation pathway: Glycosyltransferases and proteins implicated in congenital muscular dystrophy. *J. Biol. Chem.* 288:6930–6935.

Vocadlo, D. J. 2012. O-GlcNAc processing enzymes: Catalytic mechanisms, substrate specificity, and enzyme regulation. *Curr. Opin. Chem. Biol.* 16:488–497.

Hurtado-Guerrero, R., and Davies, G. J. 2012. Recent structural and mechanistic insights into post-translational enzymatic glycosylation. *Curr. Opin. Chem. Biol.* 16:479–487.

Hart, G. W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. 2011. Cross talk between O-GlcNAcylation and phosphorylation: Roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* 80:825–858.

Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* 469:564–569.

Lee, W.-S., Kang, C., Drayna, D., and Kornfeld, S. 2011. Analysis of mannose 6-phosphate uncovering enzyme mutations associated with persistent stuttering. *J. Biol. Chem.* 286:39786–39793.

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. 2008. Glycosyltransferases: Structures, functions and mechanisms. *Annu. Rev. Biochem.* 77:521–555.

Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. 2005. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem. Sci.* 30:53–62.

Carbohydrate-Binding Proteins

Gabius, H.-J., André, S., Jiménez-Barbero, J., Romero, A., and Solís, D. 2011. From lectin structure to functional glycomics: Principles of the sugar code. *Trends Biochem. Sci.* 36:298–313.

Wasserman, P. M. 2008. Zona pellucida glycoproteins. *J. Biol. Chem.* 283:24285–24289.

Sharon, N. 2008. Lectins: Past, present and future. *Biochem. Soc. Trans.* 36:1457–1460.

Balzarini, J. 2007. Targeting the glycans of glycoproteins: A novel paradigm for antiviral therapy. *Nat. Rev. Microbiol.* 5:583–597.

Sharon, N. 2007. Lectins: Carbohydrate-specific reagents and biological recognition molecules. *J. Biol. Chem.* 282:2753–2764.

Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. 2006. Structure and receptor specificity of hemagglutinin from an H5N1 influenza virus. *Science* 312:404–409.

Cambi, A., Koopman, M., and Figdor, C. G. 2005. How C-type lectins detect pathogens. *Cell. Microbiol.* 7:481–488.

Clothia, C., and Jones, E. V. 1997. The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* 66:823–862.

Bouckaert, J., Hamelryck, T., Wyns, L., and Loris, R. 1999. Novel structures of plant lectins and their complexes with carbohydrates. *Curr. Opin. Struct. Biol.* 9:572–577.

Weis, W. I., and Drickamer, K. 1996. Structural basis of lectin–carbohydrate recognition. *Annu. Rev. Biochem.* 65:441–473.

Carbohydrate Sequencing

Venkataraman, G., Shriver, Z., Raman, R., and Sasisekharan, R. 1999. Sequencing complex polysaccharides. *Science* 286:537–542.

Zhao, Y., Kent, S. B. H., and Chait, B. T. 1997. Rapid, sensitive structure analysis of oligosaccharides. *Proc. Natl. Acad. Sci. U.S.A.* 94:1629–1633.

Rudd, P. M., Guile, G. R., Küster, B., Harvey, D. J., Opdenakker, G., and Dwek, R. A. 1997. Oligosaccharide sequencing technology. *Nature* 388:205–207.

Chapter 12

Where to Start

De Weer, P. 2000. A century of thinking about cell membranes. *Annu. Rev. Physiol.* 62:919–926.

Bretscher, M. S. 1985. The molecules of the cell membrane. *Sci. Am.* 253(4):100–108.

Unwin, N., and Henderson, R. 1984. The structure of proteins in biological membranes. *Sci. Am.* 250(2):78–94.

Deisenhofer, J., and Michel, H. 1989. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis. EMBO J.* 8:2149–2170.

Singer, S. J., and Nicolson, G. L. 1972. The fluid mosaic model of the structure of cell membranes. *Science* 175:720–731.

Jacobson, K., Sheets, E. D., and Simson, R., 1995. Revisiting the fluid mosaic model of membranes. *Science* 268:1441–1442.

Books

Gennis, R. B. 1989. *Biomembranes: Molecular Structure and Function.* Springer Verlag.

Vance, D. E., and Vance, J. E. (Eds.). 2008. *Biochemistry of Lipids, Lipoproteins, and Membranes* (5th ed.). Elsevier.

Lipowsky, R., and Sackmann, E. 1995. *The Structure and Dynamics of Membranes.* Elsevier.

Racker, E. 1985. *Reconstitutions of Transporters, Receptors, and Pathological States.* Academic Press.

Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (2d ed.). Wiley-Interscience.

Membrane Lipids and Dynamics

Lingwood, D., and Simons, K. 2010. Lipid rafts as a membrane-organizing principle. *Science.* 327:46–50.

Pike, L. J. 2009. The challenge of lipid rafts. *J. Lipid Res.* 50:S323–S328.

Simons, K., and Vaz, W. L. 2004. Model systems, lipid rafts, and cell membranes. *Annu. Rev. Biophys. Biomol. Struct.* 33:269–295.

Anderson, T. G., and McConnell, H. M. 2002. A thermodynamic model for extended complexes of cholesterol and phospholipid. *Biophys. J.* 83:2039–2052.

Saxton, M. J., and Jacobson, K. 1997. Single-particle tracking: Applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* 26:373–399.

Bloom, M., Evans, E., and Mouritsen, O. G. 1991. Physical properties of the fluid lipid-bilayer component of cell membranes: A perspective. *Q. Rev. Biophys.* 24:293–397.

Elson, E. L. 1986. Membrane dynamics studied by fluorescence correlation spectroscopy and photobleaching recovery. *Soc. Gen. Physiol. Ser.* 40:367–383.

Zachowski, A., and Devaux, P. F. 1990. Transmembrane movements of lipids. *Experientia* 46:644–656.

Devaux, P. F. 1992. Protein involvement in transmembrane lipid asymmetry. *Annu. Rev. Biophys. Biomol. Struct.* 21:417–439.

Silvius, J. R. 1992. Solubilization and functional reconstitution of biomembrane components. *Annu. Rev. Biophys. Biomol. Struct.* 21:323–348.

Yeagle, P. L., Albert, A. D., Boesze-Battaglia, K., Young, J., and Frye, J. 1990. Cholesterol dynamics in membranes. *Biophys. J.* 57:413–424.

Nagle, J. F., and Tristram-Nagle, S. 2000. Lipid bilayer structure. *Curr. Opin. Struct. Biol.* 10:474–480.

Dowhan, W. 1997. Molecular basis for membrane phospholipid diversity: Why are there so many lipids? *Annu. Rev. Biochem.* 66:199–232.

Huijbregts, R. P. H., de Kroon, A. I. P. M., and de Kruijff, B. 1998. Rapid transmembrane movement of newly synthesized phosphatidylethanolamine across the inner membrane of *Escherichia coli. J. Biol. Chem.* 273:18936–18942.

Structure of Membrane Proteins

Walian, P., Cross, T. A., and Jap, B. K. 2004. Structural genomics of membrane proteins. *Genome Biol.* 5:215.

Werten, P. J., Remigy, H. W., de Groot, B. L., Fotiadis, D., Philippsen, A., Stahlberg, H., Grubmuller, H., and Engel, A. 2002. Progress in the analysis of membrane protein structure and function. *FEBS Lett.* 529:65–72.

Popot, J.-L., and Engleman, D. M. 2000. Helical membrane protein folding, stability and evolution. *Annu. Rev. Biochem.* 69:881–922.

White, S. H., and Wimley, W. C. 1999. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28:319–365.

Marassi, F. M., and Opella, S. J. 1998. NMR structural studies of membrane proteins. *Curr. Opin. Struct. Biol.* 8:640–648.

Lipowsky, R. 1991. The conformation of membranes. *Nature* 349:475–481.

Altenbach, C., Marti, T., Khorana, H. G., and Hubbell, W. L. 1990. Transmembrane protein structure: Spin labeling of bacteriorhodopsin mutants. *Science* 248:1088–1092.

Fasman, G. D., and Gilbert, W. A. 1990. The prediction of transmembrane protein sequences and their conformation: An evaluation. *Trends Biochem. Sci.* 15:89–92.

Jennings, M. L. 1989. Topography of membrane proteins. *Annu. Rev. Biochem.* 58:999–1027.

Engelman, D. M., Steitz, T. A., and Goldman, A. 1986. Identifying non-polar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15:321–353.

Udenfriend, S., and Kodukola, K. 1995. How glycosyl-phosphatidylinositol-anchored membrane proteins are made. *Annu. Rev. Biochem.* 64:563–591.

Intracellular Membranes

Skehel, J. J., and Wiley, D. C. 2000. Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annu. Rev. Biochem.* 69:531–569.

Roth, M. G. 1999. Lipid regulators of membrane traffic through the Golgi complex. *Trends Cell Biol.* 9:174–179.

Jahn, R., and Sudhof, T. C. 1999. Membrane fusion and exocytosis. *Annu. Rev. Biochem.* 68:863–911.

Stroud, R. M., and Walter, P. 1999. Signal sequence recognition and protein targeting. *Curr. Opin. Struct. Biol.* 9:754–759.

Teter, S. A., and Klionsky, D. J. 1999. How to get a folded protein across a membrane. *Trends Cell Biol.* 9:428–431.

Hettema, E. H., Distel, B., and Tabak, H. F. 1999. Import of proteins into peroxisomes. *Biochim. Biophys. Acta* 1451:17–34.

Membrane Fusion

Sollner, T. H., and Rothman, J. E. 1996. Molecular machinery mediating vesicle budding, docking and fusion. *Experientia* 52:1021–1025.
Ungar, D., and Hughson, F. M. 2003. SNARE protein structure and function. *Annu. Rev. Cell Dev. Biol.* 19:493–517.
Martens, S., and McMahon, H. T. 2008. Mechanisms of membrane fusion: Disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* 9:543–556.

Chapter 13

Where to Start

Lancaster, C. R. 2004. Structural biology: Ion pump in the movies. *Nature* 432:286–287.
Unwin, N. 2003. Structure and action of the nicotinic acetylcholine receptor explored by electron microscopy. *FEBS Lett.* 555:91–95.
Abramson, J., Smirnova, I., Kasho, V., Verner, G., Iwata, S., and Kaback, H. R. 2003. The lactose permease of *Escherichia coli:* Overall structure, the sugar-binding site and the alternating access model for transport. *FEBS Lett.* 555:96–101.
Lienhard, G. E., Slot, J. W., James, D. E., and Mueckler, M. M. 1992. How cells absorb glucose. *Sci. Am.* 266(1):86–91.
King, L. S., Kozono, D., and Agre, P. 2004. From structure to disease: The evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.* 5:687–698.
Neher, E., and Sakmann, B. 1992. The patch clamp technique. *Sci. Am.* 266(3):28–35.
Sakmann, B. 1992. Elementary steps in synaptic transmission revealed by currents through single ion channels. *Science* 256:503–512.

Books

Ashcroft, F. M. 2000. *Ion Channels and Disease.* Academic Press.
Conn, P. M. (Ed.). 1998. *Ion Channels,* vol. 293, *Methods in Enzymology.* Academic Press.
Aidley, D. J., and Stanfield, P. R. 1996. *Ion Channels: Molecules in Action.* Cambridge University Press.
Hille, B. 2001. *Ionic Channels of Excitable Membranes* (3d ed.). Sinauer.
Läuger, P. 1991. *Electrogenic Ion Pumps.* Sinauer.
Stein, W. D. 1990. *Channels, Carriers, and Pumps: An Introduction to Membrane Transport.* Academic Press.
Hodgkin, A. 1992. *Chance and Design: Reminiscences of Science in Peace and War.* Cambridge University Press.

P-Type ATPases

Sorensen, T. L., Moller, J. V., and Nissen, P. 2004. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* 304:1672–1675.
Sweadner, K. J., and Donnet, C. 2001. Structural similarities of Na, K-ATPase and SERCA, the Ca^{2+}-ATPase of the sarcoplasmic reticulum. *Biochem. J.* 356:685–704.
Toyoshima, C., and Mizutani, T. 2004. Crystal structure of the calcium pump with a bound ATP analogue. *Nature* 430:529–535.
Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405:647–655.
Auer, M., Scarborough, G. A., and Kuhlbrandt, W. 1998. Three-dimensional map of the plasma membrane H^+-ATPase in the open conformation. *Nature* 392:840–843.
Axelsen, K. B., and Palmgren, M. G. 1998. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol.* 46:84–101.
Pedersen, P. A., Jorgensen, J. R., and Jorgensen, P. L. 2000. Importance of conserved α-subunit segment ^{709}GDGVND for Mg^{2+} binding, phosphorylation, energy transduction in Na, K-ATPase. *J. Biol. Chem.* 275:37588–37595.
Blanco, G., and Mercer, R. W. 1998. Isozymes of the Na-K-ATPase: Heterogeneity in structure, diversity in function. *Am. J. Physiol.* 275:F633–F650.
Estes, J. W., and White, P. D. 1965. William Withering and the purple foxglove. *Sci. Am.* 212(6):110–117.

ATP-Binding Cassette Proteins

Locher, K. P. 2009. Structure and mechanism of ATP-binding cassette transporters. *Phil. Trans. R. Soc. B* 364:239–245.
Rees, D. C., Johnson, E., and Lewinson, O. 2009. ABC transporters: The power to change. *Nat. Rev. Mol. Cell Biol.* 10:218–227.
Ward, A., Reyes, C. L., Yu, J., Roth, C. B., and Chang, G. 2007. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci. U.S.A.* 104:19005–19010.
Locher, K. P., Lee, A. T., and Rees, D. C. 2002. The *E. coli* BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science* 296:1091–1098.
Borths, E. L., Locher, K. P., Lee, A. T., and Rees, D. C. 2002. The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proc. Natl. Acad. Sci. U.S.A.* 99:16642–16647.
Dong, J., Yang, G., and McHaourab, H. S. 2005. Structural basis of energy transduction in the transport cycle of MsbA. *Science* 308:1023–1028.
Akabas, M. H. 2000. Cystic fibrosis transmembrane conductance regulator: Structure and function of an epithelial chloride channel. *J. Biol. Chem.* 275:3729–3732.
Chen, J., Sharma, S., Quiocho, F. A., and Davidson, A. L. 2001. Trapping the transition state of an ATP-binding cassette transporter: Evidence for a concerted mechanism of maltose transport. *Proc. Natl. Acad. Sci. U.S.A.* 98:1525–1530.
Sheppard, D. N., and Welsh, M. J. 1999. Structure and function of the CFTR chloride channel. *Physiol. Rev.* 79:S23–S45.
Chen, Y., and Simon, S. M. 2000. In situ biochemical demonstration that P-glycoprotein is a drug efflux pump with broad specificity. *J. Cell Biol.* 148:863–870.
Saier, M. H., Jr., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R. A., and Nikaido, H. 1998. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *FASEB J.* 12:265–274.

Symporters and Antiporters

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. 2003. Structure and mechanism of the lactose permease of *Escherichia coli. Science* 301:610–615.
Philipson, K. D., and Nicoll, D. A. 2000. Sodium-calcium exchange: A molecular perspective. *Annu. Rev. Physiol.* 62:111–133.
Pao, S. S., Paulsen, I. T., and Saier, M. H., Jr. 1998. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* 62:1–34.
Wright, E. M., Hirsch, J. R., Loo, D. D., and Zampighi, G. A. 1997. Regulation of Na^+/glucose cotransporters. *J. Exp. Biol.* 200:287–293.
Kaback, H. R., Bibi, E., and Roepe, P. D. 1990. β-Galactoside transport in *E. coli:* A functional dissection of lac permease. *Trends Biochem. Sci.* 8:309–314.
Hilgemann, D. W., Nicoll, D. A., and Philipson, K. D. 1991. Charge movement during Na^+ translocation by native and cloned cardiac Na^+/Ca^{2+} exchanger. *Nature* 352:715–718.
Hediger, M. A., Turk, E., and Wright, E. M. 1989. Homology of the human intestinal Na^+/glucose and *Escherichia coli* Na^+/proline cotransporters. *Proc. Natl. Acad. Sci. U.S.A.* 86:5748–5752.

Ion Channels

Zhou, Y., and MacKinnon, R. 2003. The occupancy of ions in the K1 selectivity filter: Charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J. Mol. Biol.* 333:965–975.

Zhou, Y., Morais-Cabral, J. H., Kaufman, A., and MacKinnon, R. 2001. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution. *Nature* 414:43–48.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B. T., and MacKinnon, R. 2002. The open pore conformation of potassium channels. *Nature* 417:523–526.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. 2003. X-ray structure of a voltage-dependent K^+ channel. *Nature* 423:33–41.

Jiang, Y., Ruta, V., Chen, J., Lee, A., and MacKinnon, R. 2003. The principle of gating charge movement in a voltage-dependent K^+ channel. *Nature* 423:42–48.

Mackinnon, R. 2004. Structural biology: Voltage sensor meets lipid membrane. *Science* 306:1304–1305.

Noskov, S. Y., Bernèche, S., and Roux, B. 2004. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* 431:830–834.

Bezanilla, F. 2000. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* 80:555–592.

Shieh, C.-C., Coghlan, M., Sullivan, J. P., and Gopalakrishnan, M. 2000. Potassium channels: Molecular defects, diseases, and therapeutic opportunities. *Pharmacol. Rev.* 52:557–594.

Horn, R. 2000. Conversation between voltage sensors and gates of ion channels. *Biochemistry* 39:15653–15658.

Perozo, E., Cortes, D. M., and Cuello, L. G. 1999. Structural rearrangements underlying K^+-channel activation gating. *Science* 285:73–78.

Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon R. 1998. The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. *Science* 280:69–77.

Marban, E., Yamagishi, T., and Tomaselli, G. F. 1998. Structure and function of the voltage-gated Na^+ channel. *J. Physiol.* 508: 647–657.

Miller, R. J. 1992. Voltage-sensitive Ca^{2+} channels. *J. Biol. Chem.* 267: 1403–1406.

Catterall, W. A. 1991. Excitation-contraction coupling in vertebrate skeletal muscle: A tale of two calcium channels. *Cell* 64:871–874.

Ligand-Gated Ion Channels

Unwin, N. 2005. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* 346:967–989.

Miyazawa, A., Fujiyoshi, Y., Stowell, M., and Unwin, N. 1999. Nicotinic acetylcholine receptor at 4.6 Å resolution: Transverse tunnels in the channel wall. *J. Mol. Biol.* 288:765–786.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B. T., and MacKinnon, R. 2002. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* 417:515–522.

Barrantes, F. J., Antollini, S. S., Blanton, M. P., and Prieto, M. 2000. Topography of the nicotinic acetylcholine receptor membrane-embedded domains. *J. Biol. Chem.* 275:37333–37339.

Cordero-Erausquin, M., Marubio, L. M., Klink, R., and Changeux, J. P. 2000. Nicotinic receptor function: New perspectives from knockout mice. *Trends Pharmacol. Sci.* 21:211–217.

Le Novère, N., and Changeux, J. P. 1995. Molecular evolution of the nicotinic acetylcholine receptor: An example of multigene family in excitable cells. *J. Mol. Evol.* 40:155–172.

Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., and Morikawa, K. 2000. Structural basis of glutamate recognition by dimeric metabotropic glutamate receptor. *Nature* 407:971–978.

Betz, H., Kuhse, J., Schmieden, V., Laube, B., Kirsch, J., and Harvey, R. J. 1999. Structure and functions of inhibitory and excitatory glycine receptors. *Ann. N. Y. Acad. Sci.* 868:667–676.

Unwin, N. 1995. Acetylcholine receptor channel imaged in the open state. *Nature* 373:37–43.

Colquhoun, D., and Sakmann, B. 1981. Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels. *Nature* 294:464–466.

Long QT Syndrome and hERG

Saenen, J. B., and Vrints, C. J. 2008. Molecular aspects of the congenital and acquired Long QT Syndrome: Clinical implications. *J. Mol. Cell. Cardiol.* 44:633–646.

Zaręba, W. 2007. Drug induced QT prolongation. *Cardiol. J.* 14:523–533.

Fernandez, D., Ghanta, A., Kauffman, G. W., and Sanguinetti, M. C. 2004. Physicochemical features of the hERG channel drug binding site. *J. Biol. Chem.* 279:10120–10127.

Mitcheson, J. S., Chen, J., Lin, M., Culberson, C., and Sanguinetti, M. C. 2000. A structural basis for drug-induced long QT syndrome. *Proc. Natl. Acad. Sci. U.S.A.* 97:12329–12333.

Gap Junctions

Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y., and Tsukihara, T. 2009. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature* 458:597–604.

Saez, J. C., Berthoud, V. M., Branes, M. C., Martinez, A. D., and Beyer, E. C. 2003. Plasma membrane channels formed by connexins: Their regulation and functions. *Physiol. Rev.* 83:1359–1400.

Revilla, A., Bennett, M. V. L., and Barrio, L. C. 2000. Molecular determinants of membrane potential dependence in vertebrate gap junction channels. *Proc. Natl. Acad. Sci. U.S.A.* 97:14760–14765.

Unger, V. M., Kumar, N. M., Gilula, N. B., and Yeager, M. 1999. Three-dimensional structure of a recombinant gap junction membrane channel. *Science* 283:1176–1180.

Simon, A. M. 1999. Gap junctions: More roles and new structural data. *Trends Cell Biol.* 9:169–170.

Beltramello, M., Piazza, V., Bukauskas, F. F., Pozzan, T., and Mammano, F. 2005. Impaired permeability to Ins(1,4,5)P3 in a mutant connexin underlies recessive hereditary deafness. *Nat. Cell Biol.* 7:63–69.

White, T. W., and Paul, D. L. 1999. Genetic diseases and gene knockouts reveal diverse connexin functions. *Annu. Rev. Physiol.* 61:283–310.

Water Channels

Agre, P., King, L. S., Yasui, M., Guggino, W. B., Ottersen, O. P., Fujiyoshi, Y., Engel, A., and Nielsen, S. 2002. Aquaporin water channels: From atomic structure to clinical medicine. *J. Physiol.* 542:3–16.

Agre, P., and Kozono, D. 2003. Aquaporin water channels: Molecular mechanisms for human diseases. *FEBS Lett.* 555:72–78.

de Groot, B. L., Engel, A., and Grubmuller, H. 2003. The structure of the aquaporin-1 water channel: A comparison between cryo-electron microscopy and X-ray crystallography. *J. Mol. Biol.* 325:485–493.

Chapter 14

Where to Start

Scott, J. D., and Pawson, T. 2000. Cell communication: The inside story. *Sci. Am.* 282(6):7279.

Pawson, T. 1995. Protein modules and signalling networks. *Nature* 373:573–580.

Okada, T., Ernst, O. P., Palczewski, K., and Hofmann, K. P. 2001. Activation of rhodopsin: New insights from structural and biochemical studies. *Trends Biochem. Sci.* 26:318–324.

Tsien, R. Y. 1992. Intracellular signal transduction in four dimensions: From molecular design to physiology. *Am. J. Physiol.* 263: C723–C728.

Loewenstein, W. R. 1999. *Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life.* Oxford University Press.

G Proteins and 7TM Receptors

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., et al. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739–745.

Rasmussen, S. G. F., Choi, H.-J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., et al. 2007. Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* 450:383–387.

Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C., et al. 2007. GPCR engineering yields high-resolution structural insights into β_2-adrenergic receptor function. *Science* 318:1266–1273.

Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., et al. 2011. Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* 477:549–555.

Lefkowitz, R. J. 2000. The superfamily of heptahelical receptors. *Nat. Cell Biol.* 2:E133–E136.

Audet, M., and Bouvier, M. 2012. Restructuring G-protein-coupled receptor activation. *Cell* 151:14–22.

Bourne, H. R., Sanders, D. A., and McCormick, F. 1991. The GTPase superfamily: Conserved structure and molecular mechanism. *Nature* 349:117–127.

Lambright, D. G., Noel, J. P., Hamm, H. E., and Sigler, P. B. 1994. Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature* 369:621–628.

Noel, J. P., Hamm, H. E., and Sigler, P. B. 1993. The 2.2 Å crystal structure of transducin-α complexed with GTPγS. *Nature* 366: 654–663.

Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E., and Sigler, P. B. 1994. GTPase mechanism of G proteins from the 1.7-Å crystal structure of transducin α-GDP-AIF$^-_4$. *Nature* 372:276–279.

Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E., and Sigler, P. B. 1996. Crystal structure of a G-protein $\beta\gamma$ dimer at 2.1 Å resolution. *Nature* 379:369–374.

Wedegaertner, P. B., Wilson, P. T., and Bourne, H. R. 1995. Lipid modifications of trimeric G proteins. *J. Biol. Chem.* 270: 503–506.

Farfel, Z., Bourne, H. R., and Iiri, T. 1999. The expanding spectrum of G protein diseases. *New Engl. J. Med.* 340:1012–1020.

Bockaert, J., and Pin, J. P. 1999. Molecular tinkering of G protein-coupled receptors: An evolutionary success. *EMBO J.* 18: 1723–1729.

Cyclic AMP Cascade

Hurley, J. H. 1999. Structure, mechanism, and regulation of mammalian adenylyl cyclase. *J. Biol. Chem.* 274:7599–7602.

Kim, C., Xuong, N. H., and Taylor, S. S. 2005. Crystal structure of a complex between the catalytic and regulatory (RI) subunits of PKA. *Science* 307:690–696.

Tesmer, J. J., Sunahara, R. K., Gilman, A. G., and Sprang, S. R. 1997. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with G$_{s\alpha}$-GTPγS. *Science* 278:1907–1916.

Smith, C. M., Radzio-Andzelm, E., Madhusudan, Akamine, P., and Taylor, S. S. 1999. The catalytic subunit of cAMP-dependent protein kinase: Prototype for an extended network of communication. *Prog. Biophys. Mol. Biol.* 71:313–341.

Taylor, S. S., Buechler, J. A., and Yonemoto, W. 1990. cAMP-dependent protein kinase: Framework for a diverse family of regulatory enzymes. *Annu. Rev. Biochem.* 59:971–1005.

Phosphoinositide Cascade

Berridge, M. J., and Irvine, R. F. 1989. Inositol phosphates and cell signalling. *Nature* 341:197–205.

Berridge, M. J. 1993. Inositol trisphosphate and calcium signalling. *Nature* 361:315–325.

Essen, L. O., Perisic, O., Cheung, R., Katan, M., and Williams, R. L. 1996. Crystal structure of a mammalian phosphoinositide-specific phospholipase C δ. *Nature* 380:595–602.

Ferguson, K. M., Lemmon, M. A., Schlessinger, J., and Sigler, P. B. 1995. Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell* 83:1037–1046.

Baraldi, E., Carugo, K. D., Hyvonen, M., Surdo, P. L., Riley, A. M., Potter, B. V., O'Brien, R., Ladbury, J. E., and Saraste, M. 1999. Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate. *Struct. Fold. Design* 7:449–460.

Waldo, G. L., Ricks, T. K., Hicks, S. N., Cheever, M. L., Kawano, T., Tsuboi, K., Wang, X., Montell, C., Kozasa, T., Sondek, J., et al. 2010. Kinetic scaffolding mediated by a phospholipase C-β and G$_q$ signaling complex. *Science* 330:974–980.

Calcium

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* 256:632–638.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B., and Bax, A. 1995. Solution structure of calcium-free calmodulin. *Nat. Struct. Biol.* 2:768–776.

Grynkiewicz, G., Poenie, M., and Tsien, R. Y. 1985. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* 260:3440–3450.

Kerr, R., Lev-Ram, V., Baird, G., Vincent, P., Tsien, R. Y., and Schafer, W. R. 2000. Optical imaging of calcium transients in neurons and pharyngeal muscle of *C. elegans*. *Neuron* 26:583–594.

Chin, D., and Means, A. R. 2000. Calmodulin: A prototypical calcium sensor. *Trends Cell Biol.* 10:322–328.

Dawson, A. P. 1997. Calcium signalling: How do IP3 receptors work? *Curr. Biol.* 7:R544–R547.

Protein Kinases, Including Receptor Tyrosine Kinases

Riedel, H., Dull, T. J., Honegger, A. M., Schlessinger, J., and Ullrich, A. 1989. Cytoplasmic domains determine signal specificity, cellular routing characteristics and influence ligand binding of epidermal growth factor and insulin receptors. *EMBO J.* 8:2943–2954.

Taylor, S. S., Knighton, D. R., Zheng, J., Sowadski, J. M., Gibbs, C. S., and Zoller, M. J. 1993. A template for the protein kinase family. *Trends Biochem. Sci.* 18:84–89.

Sicheri, F., Moarefi, I., and Kuriyan, J. 1997. Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385:602–609.

Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., and Kuriyan, J. 1993. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: Crystal structures of the complexed and peptide-free forms. *Cell* 72:779–790.

Schlessinger, J. 2000. Cell signaling by receptor tyrosine kinases. *Cell* 103:211–225.

Simon, M. A. 2000. Receptor tyrosine kinases: Specific outcomes from general signals. *Cell* 103:13–15.

Robinson, D. R., Wu, Y. M., and Lin, S. F. 2000. The protein tyrosine kinase family of the human genome. *Oncogene* 19:5548–5557.

Hubbard, S. R. 1999. Structural analysis of receptor tyrosine kinases. *Prog. Biophys. Mol. Biol.* 71:343–358.

Carter-Su, C., and Smit, L. S. 1998. Signaling via JAK tyrosine kinases: Growth hormone receptor as a model system. *Recent Prog. Horm. Res.* 53:61–82.

Insulin Signaling Pathway

Khan, A. H., and Pessin, J. E. 2002. Insulin regulation of glucose uptake: A complex interplay of intracellular signalling pathways. *Diabetologia* 45:1475–1483.

Bevan, P. 2001. Insulin signalling. *J. Cell Sci.* 114:1429–1430.

De Meyts, P., and Whittaker, J. 2002. Structural biology of insulin and IGF1 receptors: Implications for drug design. *Nat. Rev. Drug Discov.* 1:769–783.

Dhe-Paganon, S., Ottinger, E. A., Nolte, R. T., Eck, M. J., and Shoelson, S. E. 1999. Crystal structure of the pleckstrin homology-phosphotyrosine binding (PH-PTB) targeting region of insulin receptor substrate 1. *Proc. Natl. Acad. Sci. U.S.A.* 96:8378–8383.

Domin, J., and Waterfield, M. D. 1997. Using structure to define the function of phosphoinositide 3-kinase family members. *FEBS Lett.* 410:91–95.

Hubbard, S. R. 1997. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* 16:5572–5581.

Hubbard, S. R., Wei, L., Ellis, L., and Hendrickson, W. A. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372:746–754.

EGF Signaling Pathway

Burgess, A. W., Cho, H. S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W., and Yokoyama, S. 2003. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* 12:541–552.

Cho, H. S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., Jr., and Leahy, D. J. 2003. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421:756–760.

Chong, H., Vikis, H. G., and Guan, K. L. 2003. Mechanisms of regulating the Raf kinase family. *Cell. Signal.* 15:463–469.

Stamos, J., Sliwkowski, M. X., and Eigenbrot, C. 2002. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* 277:46265–46272.

Ras

Milburn, M. V., Tong, L., deVos, A. M., Brunger, A., Yamaizumi, Z., Nishimura, S., and Kim, S. H. 1990. Molecular switch for signal transduction: Structural differences between active and inactive forms of protooncogenic Ras proteins. *Science* 247:939–945.

Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D., and Kuriyan, J. 1998. The structural basis of the activation of Ras by Sos. *Nature* 394:337–343.

Maignan, S., Guilloteau, J. P., Fromage, N., Arnoux, B., Becquart, J., and Ducruix, A. 1995. Crystal structure of the mammalian Grb2 adaptor. *Science* 268:291–293.

Takai, Y., Sasaki, T., and Matozaki, T. 2001. Small GTP-binding proteins. *Physiol. Rev.* 81:153–208.

Cancer

Druker, B. J., Sawyers, C. L., Kantarjian, H., Resta, D. J., Reese, S. F., Ford, J. M., Capdeville, R., and Talpaz, M. 2001. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *New Engl. J. Med.* 344:1038–1042.

Vogelstein, B., and Kinzler, K. W. 1993. The multistep nature of cancer. *Trends Genet.* 9:138–141.

Ellis, C. A., and Clark, G. 2000. The importance of being K-Ras. *Cell. Signal.* 12:425–434.

Hanahan, D., and Weinberg, R. A. 2000. The hallmarks of cancer. *Cell* 100:57–70.

McCormick, F. 1999. Signalling networks that cause cancer. *Trends Cell Biol.* 9:M53–M56.

Chapter 15

Where to Start

Stipanuk, M. H. (Ed.). 2006. *Biochemical, Physiological, Molecular Aspects of Human Nutrition.* Saunders-Elsevier.

McGrane, M. M., Yun, J. S., Patel, Y. M., and Hanson, R. W. 1992. Metabolic control of gene expression: In vivo studies with transgenic mice. *Trends Biochem. Sci.* 17:40–44.

Westheimer, F. H. 1987. Why nature chose phosphates. *Science* 235:1173–1178.

Kamerlin, S. C. L., Sharma, P. K., Prasad, R. B., and Warshel, A. 2013. Why nature really chose phosphate. *Q. Rev. Biophys.* 46:1–132.

Books

Atkins, P., and de Paula, J. 2011. *Physical Chemistry for the Life Sciences* (2d ed.). W. H. Freeman and Company.

Harold, F. M. 1986. *The Vital Force: A Study of Bioenergetics.* W. H. Freeman and Company.

Krebs, H. A., and Kornberg, H. L. 1957. *Energy Transformations in Living Matter.* Springer Verlag.

Nicholls, D. G., and Ferguson, S. J. 2013. *Bioenergetics* (4th ed.). Academic Press.

Frayn, K. N. 2010. *Metabolic Regulation: A Human Perspective* (3d ed.). Wiley-Blackwell.

Fell, D. 1997. *Understanding the Control of Metabolism.* Portland Press.

Harris, D. A. 1995. *Bioenergetics at a Glance.* Blackwell Scientific.

Von Baeyer, H. C. 1999. *Warmth Disperses and Time Passes: A History of Heat.* Modern Library.

Thermodynamics

Alberty, R. A. 1993. Levels of thermodynamic treatment of biochemical reaction systems. *Biophys. J.* 65:1243–1254.

Alberty, R. A., and Goldberg, R. N. 1992. Standard thermodynamic formation properties for the adenosine 5′-triphosphate series. *Biochemistry* 31:10610–10615.

Alberty, R. A. 1968. Effect of pH and metal ion concentration on the equilibrium hydrolysis of adenosine triphosphate to adenosine diphosphate. *J. Biol. Chem.* 243:1337–1343.

Goldberg, R. N. 1984. *Compiled Thermodynamic Data Sources for Aqueous and Biochemical Systems: An Annotated Bibliography (1930–1983).* National Bureau of Standards Special Publication 685, U.S. Government Printing Office.

Frey, P. A., and Arabshahi, A. 1995. Standard free energy change for the hydrolysis of the α,β-phosphoanhydride bridge in ATP. *Biochemistry* 34:11307–11310.

Bioenergetics and Metabolism

Schilling, C. H., Letscher, D., and Palsson, B. O. 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203:229–248.

DeCoursey, T. E., and Cherny, V. V. 2000. Common themes and problems of bioenergetics and voltage-gated proton channels. *Biochim. Biophys. Acta* 1458:104–119.

Giersch, C. 2000. Mathematical modelling of metabolism. *Curr. Opin. Plant Biol.* 3:249–253.

Rees, D. C., and Howard, J. B. 1999. Structural bioenergetics and energy transduction mechanisms. *J. Mol. Biol.* 293:343–350.

Regulation of Metabolism

Kemp, G. J. 2000. Studying metabolic regulation in human muscle. *Biochem. Soc. Trans.* 28:100–103.

Towle, H. C., Kaytor, E. N., and Shih, H. M. 1996. Metabolic regulation of hepatic gene expression. *Biochem. Soc. Trans.* 24:364–368.

Hofmeyr, J. H. 1995. Metabolic regulation: A control analytic perspective. *J. Bioenerg. Biomembr.* 27:479–490.

Historical Aspects

Kalckar, H. M. 1991. 50 years of biological research: From oxidative phosphorylation to energy requiring transport regulation. *Annu. Rev. Biochem.* 60:1–37.

Kalckar, H. M. (Ed.). 1969. *Biological Phosphorylations.* Prentice Hall.

Fruton, J. S. 1972. *Molecules and Life.* Wiley-Interscience.

Lipmann, F. 1971. *Wanderings of a Biochemist.* Wiley-Interscience.

Chapter 16

Where to Start

McCracken, A. N., and Edinger, A. L. 2013. Nutrient transporters: The Achilles' heel of anabolism. *Trends Endocrin. Met.* 24:200–208.

Curry, A. 2013. The milk revolution. *Nature* 500:20–22.

Bar-Even, A., Flamholz, A., Noor, E., and Milo, R. 2012. Rethinking glycolysis: On the biochemical logic of metabolic pathways. *Nature Chem. Biol.* 8:509–517.

Ward, P. S., and Thompson, C. B. 2012. Metabolic reprogramming: A cancer hallmark even Warburg did not anticipate. *Cancer Cell* 21:297–308.

Herling, A., König, M., Bulik, S., and Holzhütter, H.-G. 2011. Enzymatic features of the glucose metabolism in tumor cells. *FEBS J.* 278:2436–2459.

Lin, H. V., and Accili, D. 2011. Hormonal regulation of hepatic glucose production in health and disease. *Cell Metab.* 14:9–19.

Hirabayashi, J. 1996. On the origin of elementary hexoses. *Quart. Rev. Biol.*71:365–380.

Books and Reviews

Tong, L. 2013. Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* 70:863–891.

Frayn, K. N. 2010. *Metabolic Regulation: A Human Perspective* (3d ed.). Wiley-Blackwell.

Fell, D. 1997. *Understanding the Control of Metabolism.* Portland.

Fersht, A. 1999. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* W. H. Freeman and Company.

Poortmans, J. R. (Ed.). 2004. *Principles of Exercise Biochemistry.* Krager.

Structure of Glycolytic and Gluconeogenic Enzymes

Lietzan, A. D., and St. Maurice, M. 2013. A substrate-induced biotin binding pocket in the carboxyltransferase domain of pyruvate carboxylase. *J. Biol. Chem.* 288:19915−19925.

Banaszak, L., Mechin, I., Obmolova, G., Oldham, M., Chang, S. H., Ruiz, T., Radermacher, M., Kopperschläger, G., and Rypniewski, W. 2011. The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle. *J. Mol. Biol.* 407:284–297.

Lasso, G., Yu, L. P. C., Gil, D., Xiang, S., Tong, L., and Valle, M. 2010. Cryo-EM analysis reveals new insights into the mechanism of action of pyruvate carboxylase. *Structure* 18:1300–1310.

Ferreras, C., Hernández, E. D., Martínez-Costa, O. H., and Aragón, J. J. 2009. Subunit interactions and composition of the fructose 6-phosphate catalytic site and the fructose 2,6-bisphosphate allosteric site of mammalian phosphofructokinase. *J. Biol. Chem.* 284:9124–9131.

Hines, J. K., Chen, X., Nix, J. C., Fromm, H. J., and Honzatko. R. B. 2007. Structures of mammalian and bacterial fructose-1, 6-bisphosphatase reveal the basis for synergism in AMP/fructose-2, 6-bisphosphate inhibition. *J. Biol. Chem.* 282:36121–36131.

Ferreira-da-Silva, F., Pereira, P. J., Gales, L., Roessle, M., Svergun, D. I., Moradas-Ferreira, P., and Damas, A. M. 2006. The crystal and solution structures of glyceraldehyde-3-phosphate dehydrogenase reveal different quaternary structures. *J. Biol. Chem.* 281:33433–33440.

Kim, S.-G., Manes, N. P., El-Maghrabi, M. R., and Lee, Y.-H. 2006. Crystal structure of the hypoxia-inducible form of 6-phosphofructo-2-kinase/fructose-2,6-phosphatase (PFKFB3): A possible target for cancer therapy. *J. Biol. Chem.* 281:2939–2944.

Aleshin, A. E., Kirby, C., Liu, X., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J., and Honzatko, R. B. 2000. Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation. *J. Mol. Biol.* 296:1001–1015.

Jeffery, C. J., Bahnson, B. J., Chien, W., Ringe, D., and Petsko, G. A. 2000. Crystal structure of rabbit phosphoglucose isomerase, a glycolytic enzyme that moonlights as neuroleukin, autocrine motility factor, and differentiation mediator. *Biochemistry* 39:955–964.

Bernstein, B. E., and Hol, W. G. 1998. Crystal structures of substrates and products bound to the phosphoglycerate kinase active site reveal the catalytic mechanism. *Biochemistry* 37:4429–4436.

Rigden, D. J., Alexeev, D., Phillips, S. E. V., and Fothergill-Gilmore, L. A. 1998. The 2.3 Å X-ray crystal structure of *S. cerevisiae* phosphoglycerate mutase. *J. Mol. Biol.* 276:449–459.

Zhang, E., Brewer, J. M., Minor, W., Carreira, L. A., and Lebioda, L. 1997. Mechanism of enolase: The crystal structure of asymmetric dimer enolase-2-phospho-D-glycerate/enolase-phosphoenolpyruvate at 2.0 Å resolution. *Biochemistry* 36:12526–12534.

Hasemann, C. A., Istvan E. S., Uyeda, K., and Deisenhofer, J. 1996. The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase reveals distinct domain homologies. *Structure* 4:1017–1029.

Tari, L. W., Matte, A., Pugazhenthi, U., Goldie, H., and Delbaere, L. T. J. 1996. Snapshot of an enzyme reaction intermediate in the structure of the ATP-Mg^{2+}-oxalate ternary complex of *Escherichia coli* PEP carboxykinase. *Nat. Struct. Biol.* 3:355–363.

Catalytic Mechanisms

Soukri, A., Mougin, A., Corbier, C., Wonacott, A., Branlant, C., and Branlant, G. 1989. Role of the histidine 176 residue in glyceraldehyde-3-phosphate dehydrogenase as probed by site-directed mutagenesis. *Biochemistry* 28:2586–2592.

Bash, P. A., Field, M. J., Davenport, R. C., Petsko, G. A., Ringe, D., and Karplus, M. 1991. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* 30:5826–5832.

Knowles, J. R., and Albery, W. J. 1977. Perfection in enzyme catalysis: The energetics of triosephosphate isomerase. *Acc. Chem. Res.* 10: 105–111.

Regulation

Liu, S., Ammirati, M. J., Song, X., Knafels, J. D., Zhang, J., Greasley, S. E., Pfefferkorn, J. A., and Qiu, X. 2012. Insights into mechanism of glucokinase activation: Observation of multiple distinct protein conformations. *J. Biol. Chem.* 287:13598–13610.

Brüser, A., Kirchberger, J., Kloos, M., Sträter, N., and Schöneberg, T. 2012. Functional linkage of adenine nucleotide binding sites in mammalian muscle 6-phosphofructokinase. *J. Biol. Chem.* 287: 17546–17553.

Anderka, O., Boyken, J., Aschenbach, U., Batzer, A., Boscheinen, O., and Schmoll, D. 2008. Biophysical characterization of the interaction between hepatic glucokinase and its regulatory protein: Impact of physiological and pharmacological effectors. *J. Biol. Chem.* 283:31333–31340.

Iancu, C. V., Mukund, S., Fromm, H. J., and Honzatko, R. B. 2005. R-state AMP complex reveals initial steps of the quaternary transition of fructose-1,6-bisphosphatase. *J. Biol. Chem.* 280: 19737–19745.

Lee, Y. H., Li, Y., Uyeda, K., and Hasemann, C. A. 2003. Tissue-specific structure/function differentiation of the five isoforms of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. *J. Biol. Chem.* 278:523–530.

Gleeson, T. T. 1996. Post-exercise lactate metabolism: A comparative review of sites, pathways, and regulation. *Annu. Rev. Physiol.* 58:556–581.

Jitrapakdee, S., and Wallace, J. C. 1999. Structure, function and regulation of pyruvate carboxylase. *Biochem. J.* 340:1–16.

van de Werve, G., Lange, A., Newgard, C., Mechin, M. C., Li, Y., and Berteloot, A. 2000. New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. *Eur. J. Biochem.* 267:1533–1549.

Sugar Transporters

Blodgett, D. M., Graybill, C. and Carruthers, A. 2008. Analysis of glucose transporter topology and structural dynamics. *J. Biol. Chem.* 283:36416–36424.

Huang, S., and Czech, M. P. 2007. The GLUT4 glucose transporter. *Cell Metab.* 5:237–252.

Czech, M. P., and Corvera, S. 1999. Signaling mechanisms that regulate glucose transport. *J Biol. Chem.* 274:1865–1868.

Silverman, M. 1991. Structure and function of hexose transporters. *Annu. Rev. Biochem.* 60:757–794.

Thorens, B., Charron, M. J., and Lodish, H. F. 1990. Molecular physiology of glucose transporters. *Diabetes Care* 13:209–218.

Glycolysis and Cancer

Morgan, H. P., O'Reilly, F. J., Wear, M. A., O'Neill, J. R., Fothergill-Gilmore, L. A., Hupp, T., and Walkinshaw, M. D. 2013. M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. *Proc. Natl. Acad. Sci. U.S.A.* 110: 5881–5886.

Schulze, A., and Harris, A. L. 2012. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature* 491:364–373.

Lunt, S. Y., and Vander Heiden, M. G. 2011. Aerobic glycolysis: Meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* 27:441–64.

Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. 2009. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science* 324:1029–1033.

Mathupala, S. P., Ko, Y. H., and Pedersen, P. L. 2009. Hexokinase-2 bound to mitochondria: Cancer's stygian link to the "Warburg effect" and a pivotal target for effective therapy. *Sem. Cancer Biol.* 19:17–24.

Kroemer, G. K., and Pouyssegur, J. 2008. Tumor cell metabolism: Cancer's Achilles' heel. *Cancer Cell* 12:472–482.

Hsu, P. P., and Sabatini, D. M. 2008. Cancer cell metabolism: Warburg and beyond. *Cell* 134:703–707.

Genetic Diseases

Orosz, F., Oláh, J., and Ovádi, J. 2009. Triosephosphate isomerase deficiency: New insights into an enigmatic disease. *Biochim. Biophys. Acta* 1792:1168–1174.

Scriver, C. R., Beaudet, A. L., Valle, D., Sly, W. S., Childs, B., Kinzler, K., and Vogelstein, B. (Eds.). 2001. *The Metabolic and Molecular Basis of Inherited Disease* (8th ed.). McGraw-Hill.

Evolution

Dandekar, T., Schuster, S., Snel, B., Huynen, M., and Bork, P. 1999. Pathway alignment: Application to the comparative analysis of glycolytic enzymes. *Biochem. J.* 343:115–124.

Heinrich, R., Melendez-Hevia, E., Montero, F., Nuno, J. C., Stephani, A., and Waddell, T. G. 1999. The structural design of glycolysis: An evolutionary approach. *Biochem. Soc. Trans.* 27:294–298.

Walmsley, A. R., Barrett, M. P., Bringaud, F., and Gould, G. W. 1998. Sugar transporters from bacteria, parasites and mammals: Structure-activity relationships. *Trends Biochem. Sci.* 23: 476–480.

Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Thi, M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R., et al. 1999. The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima:* A comparative thermostability structural analysis of ten different TIM structures. *Proteins* 37:441–453.

Historical Aspects

Friedmann, H. C. 2004. From *Butyribacterium* to *E. coli:* An essay on unity in biochemistry. *Perspect. Biol. Med.* 47:47–66.

Fruton, J. S. 1999. *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology.* Yale University Press.

Kalckar, H. M. (Ed.). 1969. *Biological Phosphorylations: Development of Concepts*. Prentice Hall.

Chapter 17

Where to Start

Sugden, M. C., and Holness, M. J. 2003. Recent advances in mechanisms regulating glucose oxidation at the level of the pyruvate dehydrogenase complex by PDKs. *Am. J. Physiol. Endocrinol. Metab.* 284:E855–E862.

Owen, O. E., Kalhan, S. C., and Hanson, R. W. 2002. The key role of anaplerosis and cataplerosis for citric acid function. *J. Biol. Chem.* 277:30409–30412.

Pyruvate Dehydrogenase Complex

Patel, K. P., O'Brien, T. W., Subramony, S. H., Shuster, J., and Stacpoole, P. W. 2012. The spectrum of pyruvate dehydrogenase complex deficiency: Clinical, biochemical and genetic features in 371 patients. *Mol. Genet. Metab.* 105:34–43.

Vijayakrishnan, S., Callow, P., Nutley, M. A., Mcgow, D. P., Gilbert, D., Kropholler, P., Cooper, A., Byron, O., And Lindsay, J. G. 2011. Variation in the organization and subunit composition of the mammalian pyruvate dehydrogenase complex E2/E3BP core assembly. *Biochem. J.* 437:565–574.

Vijayakrishnan, S., Kelly, S. M., Gilbert, R. J., Callow, P., Bhella, D., Forsyth, T., Lindsay, J. G., and Byron, O. 2010. Solution structure and characterization of the human pyruvate dehydrogenase complex core assembly. *J. Mol. Biol.* 399:71–93.

Brautigam, C. A., Wynn, R. M., Chuang, J. L., and Chuang, D. T. 2009. Subunit and catalytic component stoichiometries of an in vitro reconstituted human pyruvate dehydrogenase complex. *J. Biol. Chem.* 284:13086–13098.

Hiromasa, Y., Fujisawa, T., Aso, Y., and Roche, T. E. 2004. Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding proteins and their capacities to bind the E1 and E3 components. *J. Biol Chem.* 279:6921–6933.

Domingo, G. J., Chauhan, H. J., Lessard, I. A., Fuller, C., and Perham, R. N. 1999. Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus. Eur. J. Biochem.* 266:1136–1146.

Structure of Citric Acid Cycle Enzymes

Fraser, M. E., Hayakawa, K., Hume, M. S., Ryan, D. G., and Brownie, E. R. 2006. Interactions of GTP with the ATP-grasp domain of GTP-specific succinyl-CoA synthetase. *J. Biol. Chem.* 281:11058–11065.

Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C., Byrne, B., Cecchini, G., and Iowata, S. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* 299:700–704.

Fraser, M. E., James, M. N., Bridger, W. A., and Wolodko, W. T. 1999. A detailed structural description of *Escherichia coli* succinyl-CoA synthetase. *J. Mol. Biol.* 285:1633–1653. [Published erratum appears in May 7, 1999, issue of *J. Mol. Biol.* 288(3):501.]

Lloyd, S. J., Lauble, H., Prasad, G. S., and Stout, C. D. 1999. The mechanism of aconitase: 1.8 Å resolution crystal structure of the S642A:citrate complex. *Protein Sci.* 8:2655–2662.

Rose, I. A. 1998. How fumarase recycles after the malate → fumarate reaction: Insights into the reaction mechanism. *Biochemistry* 37:17651–17658.

Organization of the Citric Acid Cycle

Lambeth, D. O., Tews, K. N., Adkins, S., Frohlich, D., and Milavetz, B. I. 2004. Expression of two succinyl-CoA specificities in mammalian tissues. *J. Biol. Chem.* 279:36621–36624.

Velot, C., Mixon, M. B., Teige, M., and Srere, P. A. 1997. Model of a quinary structure between Krebs TCA cycle enzymes: A model for the metabolon. *Biochemistry* 36:14271–14276.

Haggie, P. M., and Brindle, K. M. 1999. Mitochondrial citrate synthase is immobilized in vivo. *J. Biol. Chem.* 274:3941–3945.

Morgunov, I., and Srere, P. A. 1998. Interaction between citrate synthase and malate dehydrogenase: Substrate channeling of oxaloacetate. *J. Biol. Chem.* 273:29540–29544.

Regulation

Shi, Q., Xu, H., Yu, H., Zhang, N., Ye, Y., Estevez, A. G., Deng, H., and Gibson, G. E. 2011. Inactivation and reactivation of the mitochondrial α-ketoglutarate dehydrogenase complex. *J. Biol. Chem.* 286:17640–17648.

Phillips, D., Aponte, A. M., French, S. A., Chess, D. J., and Balaban, R. S. 2009. Succinyl-CoA synthetase is a phosphate target for the activation of mitochondrial metabolism. *Biochemistry* 48: 7140–7149.

Taylor, A. B., Hu, G., Hart, P. J., and McAlister-Henn, L. 2008. Allosteric motions in structures of yeast NAD$^+$-specific isocitrate dehydrogenase. *J. Biol. Chem.* 283:10872–10880.

Green, T., Grigorian, A., Klyuyeva, A., Tuganova, A., Luo, M., and Popov, K. M. 2008. Structural and functional insights into the molecular mechanisms responsible for the regulation of pyruvate dehydrogenase kinase. *J. Biol. Chem.* 283:15789–15798.

Hiromasa, Y., and Roche, T. E. 2003. Facilitated interaction between the pyruvate dehydrogenase kinase isoform 2 and the dihydrolipoyl acetyltransferases. *J. Biol. Chem.* 278:33681–33693.

Jitrapakdee, S., and Wallace, J. C. 1999. Structure, function and regulation of pyruvate carboxylase. *Biochem. J.* 340:1–16.

The Citric Acid Cycle and Cancer

Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernytsky, A., Liu, W., Gliser, C., et al. 2013. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* 340:622–626.

Rohle, D., Popovici-Muller, J., Palaskas, N., Turcan, S., Grommes, C., Campos, C., Tsoi, J., Clark, O., Oldrini, B., Komisopoulou, E., et al. 2013. An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* 340:626–630.

Losman, J.-A., Koivunen, P., Lee, S., Schneider, R. K., McMahon, C., Cowley, G. S., Root, D. E., Ebert, B. L., Kaelin, W. G. Jr., et al. 2013. (*R*)-2-Hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* 339:1621–1625.

Sakai, C., Tomitsuka, T., Esumi, H., Harada, S., and Kita, K. 2012. Mitochondrial fumarate reductase as a target of chemotherapy: From parasites to cancer cells. *Biochim. Biophys. Acta* 1820:643–651.

Xekouki P., and Stratakis, C. A. 2012. Succinate dehydrogenase (SDHx) mutations in pituitary tumors: Could this be a new role for mitochondrial complex II and/or Krebs cycle defects? *Endocr.-Relat. Cancer* 19:C33–C40.

Thompson, C. B. 2009. Metabolic enzymes as oncogenes or tumor suppressors. *New Engl. J. Med.* 360:813–815.

McFate, T., Mohyeldin, A., Lu, H., Thakar, J., Henriques, J., Halim, N. D., Wu, H., Schell, M. J., Tsang, T. M., Teahan, O., Zhou, S., Califano, J. A., Jeoung, M. N., Harris, R. A., and Verma, A. 2008. Pyruvate dehydrogenase complex activity controls metabolic and malignant phenotype in cancer cells. *J. Biol. Chem.* 283:22700–22708.

Gogvadze, V., Orrenius, S., and Zhivotovsky, B. 2008. Mitochondria in cancer cells: What is so special about them? *Trends Cell Biol.* 18:165–173.

Evolutionary Aspects

Meléndez-Hevia, E., Waddell, T. G., and Cascante, M. 1996. The puzzle of the Krebs citric acid cycle: Assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways in evolution. *J. Mol. Evol.* 43:293–303.

Baldwin, J. E., and Krebs, H. 1981. The evolution of metabolic cycles. *Nature* 291:381–382.

Gest, H. 1987. Evolutionary roots of the citric acid cycle in prokaryotes. *Biochem. Soc. Symp.* 54:3–16.

Weitzman, P. D. J. 1981. Unity and diversity in some bacterial citric acid cycle enzymes. *Adv. Microbiol. Physiol.* 22:185–244.

Discovery of the Citric Acid Cycle

Kornberg, H. 2000. Krebs and his trinity of cycles. *Nat. Rev. Mol. Cell. Biol.* 1:225–228.

Krebs, H. A., and Johnson, W. A. 1937. The role of citric acid in intermediate metabolism in animal tissues. *Enzymologia* 4:148–156.

Krebs, H. A. 1970. The history of the tricarboxylic acid cycle. *Perspect. Biol. Med.* 14:154–170.

Krebs, H. A., and Martin, A. 1981. *Reminiscences and Reflections.* Clarendon Press.

Chapter 18

Where to Start

Guarente, L. 2008. Mitochondria: A nexus for aging, calorie restriction, and sirtuins? *Cell* 132:171–176.

Wallace, D. C. 2007. Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu. Rev. Biochem.* 76:781–821.

Hosler, J. P., Ferguson-Miller, S., and Mills, D. A. 2006. Energy transduction: Proton transfer through the respiratory complexes. *Annu. Rev. Biochem.* 75:165–187.

Gray, M. W., Burger, G., and Lang, B. F. 1999. Mitochondrial evolution. *Science* 283:1476–1481.

Shultz, B. E., and Chan, S. I. 2001. Structures and proton-pumping strategies of mitochondrial respiratory enzymes. *Annu. Rev. Biophys. Biomol. Struct.* 30:23–65.

Books

Scheffler, I. E. 2007. *Mitochondria.* Wiley.

Lane, N. 2005. *Power, Sex, Suicide: Mitochondria and the Meaning of Life.* Oxford.

Nicholls, D. G., and Ferguson, S. J. 2013. *Bioenergetics* (4th ed.). Academic Press.

Electron-Transport Chain

Baradaran, R., Berrisford, J. M., Minhas, G. S., and Sazanov, L. A. 2013. Crystal structure of the entire respiratory complex I. *Nature* 494:443–448.

Lapuente-Brun, E., Moreno-Loshuertos, R., Acín-Pérez, R., Latorre-Pellicer, A., Colás, C., Balsa, E., Perales-Clemente, E., Quirós, P. M., Calvo, E., Rodríguez-Hernández, M. A., et al. 2013. Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. *Science* 340:1567–1570.

Cammack, R. 2012. Iron-sulfur proteins. *The Biochemist* 35:14–17.

Yoshikawa, S., Muramoto, K., and Shinzawa-Itoh, K. 2011. Proton-pumping mechanism of cytochrome *c* oxidase. *Annu. Rev. Biophys.* 40:205–23.

Qin, L., Liu, J., Mills, D. A., Proshlyakov, D. A., Hiser, C., and Ferguson-Miller, S. 2009. Redox-dependent conformational changes in cytochrome *c* oxidase suggest a gating mechanism for proton uptake. *Biochemistry* 48:5121–5130.

Lill, R. 2009. Function and biogenesis of iron–sulphur proteins. *Nature* 460:831–838.

Cooley, C. W., Lee, D.-W., and Daldal, F. 2009. Across membrane communication between the Q_o and Q_i active sites of cytochrome bc_1. *Biochemistry* 48:1888–1899.

Verkhovskaya, M. L., Belevich, N., Euro, L., Wikström, M., and. Verkhovsky, M. I. 2008. Real-time electron transfer in respiratory complex I. *Proc. Natl. Acad. Sci. U.S.A.* 105:3763–3767.

Acín-Pérez, R., Fernández-Silva, P., Peleato, M. L., Pérez-Martos, A., and Enriquez, J. A. 2008. Respiratory active mitochondrial supercomplexes. *Mol. Cell* 32:529–539.

Kruse, S. E., Watt, W. C., Marcinek, D. J., Kapur, R. P., Schenkman, K. A., and Palmiter, R. D. 2008. Mice with mitochondrial Complex I deficiency develop a fatal encephalomyopathy. *Cell Metab.* 7:312–320.

Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., Bartlam, M., and Ral, Z. 2005. Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* 121:1043–1057.

Crofts, A. R. 2004. The cytochrome bc_1 complex: Function in the context of structure. *Annu. Rev. Physio*l. 66:689–733.

Bianchi, C., Genova, M. L., Castelli, G. P., and Lenaz, G. 2004. The mitochondrial respiratory chain is partially organized in a supramolecular complex. *J. Biol. Chem.* 279:36562–36569.

Cecchini, G. 2003. Function and structure of Complex II of the respiratory chain. *Annu. Rev. Biochem.* 72:77–109.

Lange, C., and Hunte, C. 2002. Crystal structure of the yeast cytochrome bc_1 complex with its bound substrate cytochrome *c*. *Proc. Natl. Acad. Sci. U.S.A.* 99:2800–2805.

ATP Synthase

Toei, M., and Noji, H. 2013. Single-molecule analysis of F_0F_1-ATP synthase inhibited by N, N-dicyclohexylcarbodiimide. *J. Biol. Chem.* 288:25717–25726.

Watt, I. N., Montgomery, M. G., Runswick, M. J., Leslie, A. G. W., and Walker, J. E. 2010. Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 107:16823–16827.

Wittig, I., and Hermann, S. 2009. Supramolecular organization of ATP synthase and respiratory chain in mitochondrial membranes. *Biochim. Biophys. Acta* 1787:672–680.

Junge, W., Sielaff, H., and Engelbrecht S. 2009. Torque generation and elastic power transmission in the rotary F_0F_1-ATPase. *Nature* 459:364–370.

von Ballmoos, C., Cook, G. M., and Dimroth, P. 2008. Unique rotary ATP synthase and its biological diversity. *Annu. Rev. Biophys.* 37:43–64.

Adachi, K., Oiwa, K., Nishizaka, T., Furuike, S., Noji, H., Itoh, H., Yoshida, M., and Kinosita, K., Jr. 2007. Coupling of rotation and catalysis in F_1-ATPase revealed by single-molecule imaging and manipulation. *Cell* 130:309–321.

Chen, C., Ko, Y., Delannoy, M., Ludtke, S. J., Chiu, W., and Pedersen, P. L. 2004. Mitochondrial ATP synthasome: Three-dimensional structure by electron microscopy of the ATP synthase in complex formation with the carriers for P_i and ADP/ATP. *J. Biol. Chem.* 279:31761–31768.

Noji, H., and Yoshida, M. 2001. The rotary machine in the cell: ATP synthase. *J. Biol. Chem.* 276:1665–1668.

Yasuda, R., Noji, H., Kinosita, K., Jr., and Yoshida, M. 1998. F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* 93:1117–1124.

Noji, H., Yasuda, R., Yoshida, M., and Kinosita, K., Jr., 1997. Direct observation of the rotation of F_1-ATPase. *Nature* 386:299–302.

Tsunoda, S. P., Aggeler, R., Yoshida, M., and Capaldi, R. A. 2001. Rotation of the *c* subunit oligomer in fully functional F_1 F_0 ATP synthase. *Proc. Natl. Acad. Sci. U.S.A.* 987:898–902.

Gibbons, C., Montgomery, M. G., Leslie, A. G. W., and Walker, J. 2000. The structure of the central stalk in F_1-ATPase at 2.4 Å resolution. *Nat. Struct. Biol.* 7:1055–1061.

Sambongi, Y., Iko, Y., Tanabe, M., Omote, H., Iwamoto-Kihara, A., Ueda, I., Yanagida, T., Wada, Y., and Futai, M. 1999. Mechanical rotation of the *c* subunit oligomer in ATP synthase (F_0F_1): Direct observation. *Science* 286:1722–1724.

Translocators and Channels

Villarroya, F., and Vidal-Puig, A. 2013. Beyond the sympathetic tone: The new brown fat activators. *Cell Metab.*17:638–643.

Rey, M., Forest, E., and Pelosi, L. 2012. Exploring the conformational dynamics of the bovine ADP/ATP carrier in mitochondria. *Biochemistry* 51:9727–9735.

Divakaruni, A. S., Humphrey, D. M., and Brand, M. D. 2012. Fatty acids change the conformation of uncoupling protein 1 (UCP1). *J. Biol. Chem.* 44:36845–36853.

Fedorenko, A., Lishko, P. V., and Kirichok, Y. 2012. Mechanism of fatty-acid-dependent UCP1 uncoupling in brown fat mitochondria. *Cell* 151:400–413.

van Marken Lichtenbelt, W. D., Vanhommerig, J. W., Smulders, N. M., Drossaerts, J. M., Kemerink, G. J., Bouvy, N. D., Schrauwen, P., and Teule, G. J. 2009. Cold-activated brown adipose tissue in healthy men. *New Engl. J. Med.* 360:1500–1508.

Cypess, A. M., Sanaz Lehman, S., Gethin Williams, G., Tal, I., Rodman, D., Goldfine, A. B., Kuo, F. C., Palmer, E. L., Tseng, Y.-H., Doria, A., et al. 2009. Identification and importance of brown adipose tissue in adult humans. *New Engl. J. Med.* 360:1509–1517.

Virtanen, K. A., Lidell, M. E., Orava, J., Heglind, M., Westergren, R., Niemi, T., Taittonen, M., Laine, J., Savisto, N.-J., Enerbäck, S., et al. 2009. Functional brown adipose tissue in healthy adults. *New Engl. J. Med.* 360:1518–1525.

Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vonrhein, C., Griesinger, C., Zweckstetter, M., and Zeth, K. 2008. Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U.S.A.* 105:15370–15375.

Bamber, L., Harding, M., Monné, M., Slotboom, D.-J., and Kunji, E. R. 2007. The yeast mitochondrial ADP/ATP carrier functions as a monomer in mitochondrial membranes. *Proc. Natl. Acad. Sci. U.S.A.* 10:10830–10843.

Pebay-Peyroula, E., Dahout, C., Kahn, R., Trézéguet, V., Lauquin, G. J.-M., and Brandolin, G. 2003. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* 246:39–44.

Reactive Oxygen Species, Superoxide Dismutase, and Catalase

Sena, L. A., and Chandel, N. S. 2012. Physiological roles of mitochondrial reactive oxygen species. *Mol. Cell* 48:158–167.

Forman, H. J., Maiorino, M., and Ursini, F. 2010. Signaling functions of reactive oxygen species. *Biochemistry* 49:835–842.

Murphy, M. P. 2009. How mitochondria produce reactive oxygen species. *Biochem. J.* 417:1–13.

Leitch, J. M., Yick, P. J., and Culotta, V. V. 2009. The right to choose: Multiple pathways for activating copper, zinc superoxide dismutase. *J. Biol. Chem.* 284:24679–24683.

Winterbourn, C. C. 2008. Reconciling the chemistry and biology of reactive oxygen species. *Nat. Chem. Biol.* 4:278–286.

Veal, E. A., Day, A. M., and Morgan, B. A. 2007. Hydrogen peroxide sensing and signaling. *Mol. Cell* 26:1–14.

Stone, J. R., and Yang, S. 2006. Hydrogen peroxide: A signaling messenger. *Antioxid. Redox Signal.* 8:243–270.

Valentine, J. S., Doucette, P. A., and Potter S. Z. 2005. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annu. Rev. Biochem.* 74:563–593.

Mitochondrial Diseases

Papa, S., and De Rasmo, D. 2013. Complex I deficiencies in neurological disorders. *Trends Mol. Med.* 19:61–69.

Koopman, W. J. H., Willems, P. H. G. M., and Smeitink, J. A. M. 2012. Monogenic mitochondrial disorders. *New Engl. J. Med.* 366:1132–41.

Lina, C. S., Sharpley, M. S., Fan W., Waymire, K. G., Sadun, A. A., Carelli, V., Ross-Cisneros, F. N., Baciu, P., Sung, E., McManus, M. J., et al. 2012. Mouse mtDNA mutant model of Leber hereditary optic neuropathy. *Proc. Natl. Acad. Sci. U.S.A.* 109:20065–20070.

Mitochondria Disease. 2009. A compendium of nine articles on mitochondrial diseases. *Biochem. Biophys. Acta Mol. Basis Disease* 1792:1095–1167.

Cicchetti, F., Drouin-Ouellet, J., and Gross, R. E. 2009. Environmental toxins and Parkinson's disease: What have we learned from pesticide-induced animal models? *Trends Pharm. Sci.* 30:475–483.

DiMauro, S., and Schon, E. A. 2003. Mitochondrial respiratory-chain disease. *New Engl. J. Med.* 348:2656–2668.

Smeitink, J., van den Heuvel, L., and DiMauro, S. 2001. The genetics and pathology of oxidative phosphorylation. *Nat. Rev. Genet.* 2:342–352.

Apoptosis

Qi, S., Pang, Y., Hu, Q., Liu, Q., Li, H., Zhou, Y., He, T., Liang, Q., Liu, Y., Yuan, X., et al. 2010. Crystal structure of the *Caenorhabditis elegans* apoptosome reveals an octameric assembly of CED-4. *Cell* 141:446–457.

Chan, D. C. 2006. Mitochondria: Dynamic organelles in disease, aging, and development. *Cell* 125:1241–1252.

Green, D. R. 2005. Apoptotic pathways: Ten minutes to dead. *Cell* 121:671–674.

Historical Aspects

Prebble, J., and Weber, B. 2003. *Wandering in the Gardens of the Mind: Peter Mitchell and the Making of Glynn.* Oxford.

Mitchell, P. 1979. Keilin's respiratory chain concept and its chemiosmotic consequences. *Science* 206:1148–1159.

Preeble, J. 2002. Peter Mitchell and the ox phos wars. *Trends Biochem. Sci.* 27:209–212.

Mitchell, P. 1976. Vectorial chemistry and the molecular mechanics of chemiosmotic coupling: Power transmission by proticity. *Biochem. Soc. Trans.* 4:399–430.

Racker, E. 1980. From Pasteur to Mitchell: A hundred years of bioenergetics. *Fed. Proc.* 39:210–215.

Kalckar, H. M. 1991. Fifty years of biological research: From oxidative phosphorylation to energy requiring transport and regulation. *Annu. Rev. Biochem.* 60:1–37.

Chapter 19

Where to Start

Huber, R. 1989. A structural basis of light energy and electron transfer in biology. *EMBO J.* 8:2125–2147.

Deisenhofer, J., and Michel, H. 1989. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis. EMBO J.* 8:2149–2170.

Barber, J., and Andersson, B. 1994. Revealing the blueprint of photosynthesis. *Nature* 370:31–34.

Books and General Reviews

Nelson, N., and Yocum, C. 2006. Structure and functions of photosystems I and II. *Annu. Rev. Plant Biol.* 57:521–565.

Merchant, S., and Sawaya, M. R. 2005. The light reactions: A guide to recent acquisitions for the picture gallery. *Plant Cell* 17:648–663.

Blankenship, R. E. 2009. *Molecular Mechanisms of Photosynthesis.* Wiley-Blackwell.

Nicholls, D. G., and Ferguson, S. J. 2013. *Bioenergetics* (4th ed.). Academic Press.

Electron-Transfer Mechanisms

Beratan, D., and Skourtis, S. 1998. Electron transfer mechanisms. *Curr. Opin. Chem. Biol.* 2:235–243.

Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S., and Dutton, P. L. 1992. Nature of biological electron transfer. *Nature* 355:796–802.

Boxer, S. G. 1990. Mechanisms of long-distance electron transfer in proteins: Lessons from photosynthetic reaction centers. *Annu. Rev. Biophys. Biophys. Chem.* 19:267–299.

Photosystem II

Vinyard, D. J., Ananyev, G. M., and Dismukes, G. C. 2013. Photosystem II: The reaction center of oxygenic photosynthesis. *Annu. Rev. Biochem.* 82:577–606.

Kirchhoff, H., Tremmel, I., Haase, W., and Kubitscheck, U. 2004. Supramolecular photosystem II organization in grana of thylakoid membranes: Evidence for a structured arrangement. *Biochemistry* 43:9204–9213.

Diner, B. A., and Rappaport, F. 2002. Structure, dynamics, and energetics of the primary photochemistry of photosystem II of oxygenic photosynthesis. *Annu. Rev. Plant Biol.* 54:551–580.

Zouni, A., Witt, H. T., Kern, J., Fromme, P., Krauss, N., Saenger, W., and Orth, P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409:739–743.

Deisenhofer, J., and Michel, H. 1991. High-resolution structures of photosynthetic reaction centers. *Annu. Rev. Biophys. Biophys. Chem.* 20:247–266.

Oxygen Evolution

Umena, Y., Kawakami, K., Shen, J.-R. and Kamiya, N. 2011. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9Å. *Nature* 473:55–60.

Barber, J. 2008. Crystal structure of the oxygen-evolving complex of photosystem II. *Inorg. Chem.* 47:1700–1710.

Pushkar, Y., Yano, J., Sauer, K., Boussac, A., and Yachandra, V. K. 2008. Structural changes in the Mn_4Ca cluster and the mechanism of photosynthetic water splitting. *Proc. Natl. Acad. Sci. U.S.A.* 105:1879–1884.

Renger, G. 2007. Oxidative photosynthetic water splitting: Energetics, kinetics and mechanism. *Photosynth. Res.* 92:407–425.

Renger, G., and Kühn, P. 2007. Reaction pattern and mechanism of light induced oxidative water splitting in photosynthesis. *Biochim. Biophys. Acta* 1767:458–471.

Photosystem I and Cytochrome *bf*

Schöttler, M. A., Albus, C. A., and Bock, R. 2011. Photosystem I: Its biogenesis and function in higher plants. *J. Plant Physiol.* 168:1452–1461.

Iwai, M., Takizawa, K., Tokutsu, R., Okamuro, A., Takahashi, Y., and Minagawa, J. 2010. Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. *Nature* 464:1210–1214.

Amunts, A., Drory, O., and Nelson, N. 2007. The structure of photosystem I supercomplex at 3.4 Å resolution. *Nature* 447:58–63.

Cramer, W. A., Zhang, H., Yan, J., Kurisu, G., and Smith, J. L. 2004. Evolution of photosynthesis: Time-independent structure of the cytochrome b_6f complex. *Biochemistry* 43:5921–5929.

Kargul, J., Nield, J., and Barber, J. 2003. Three-dimensional reconstruction of a light-harvesting complex I-photosystem I (LHCI-PSI) supercomplex from the green alga *Chlamydomonas reinhardtii. J. Biol. Chem.* 278:16135–16141.

Schubert, W. D., Klukas, O., Saenger, W., Witt, H. T., Fromme, P., and Krauss, N. 1998. A common ancestor for oxygenic and anoxygenic photosynthetic systems: A comparison based on the structural model of photosystem I. *J. Mol. Biol.* 280:297–314.

ATP Synthase

Kohzuma, K., Dal Bosco, C., Meurer, J., and Kramer, D. M. 2013. Light- and metabolism-related regulation of the chloroplast ATP synthase has distinct mechanisms and functions. *J. Biol. Chem.* 288:13156–13163.

Vollmar, M., Schlieper, D., Winn, D., Büchner, C., and Groth, G. 2009. Structure of the c14 rotor ring of the proton translocating chloroplast ATP synthase. *J. Biol. Chem.* 284:18228–18235.

Varco-Merth, B., Fromme, R., Wang, M., and Fromme, P. 2008. Crystallization of the c14-rotor of the chloroplast ATP synthase reveals that it contains pigments. *Biochim. Biophys. Acta* 1777:605–612.

Oster, G., and Wang, H. 1999. ATP synthase: Two motors, two fuels. *Structure* 7:R67–R72.

Weber, J., and Senior, A. E. 2000. ATP synthase: What we know about ATP hydrolysis and what we do not know about ATP synthesis. *Biochim. Biophys. Acta* 1458:300–309.

Light-Harvesting Assemblies

Collins, A. M., Qian, P., Tang, Q., Bocian, D. F., Hunter, C. N., Blankenship, R. E. 2010. Light-harvesting antenna system from the phototrophic bacterium *Roseiflexus castenholzii. Biochemistry* 49:7524–7531.

Melkozernov, A. N., Barber, J., and Blankenship, R. E. 2006. Light harvesting in photosystem I supercomplexes. *Biochemistry* 45:331–345.

Conroy, M. J., Westerhuis, W. H., Parkes-Loach, P. S., Loach, P. A., Hunter, C. N., and Williamson, M. P. 2000. The solution structure of *Rhodobacter sphaeroides* LH1b reveals two helical domains separated by a more flexible region: Structural consequences for the LH1 complex. *J. Mol. Biol.* 298:83–94.

Koepke, J., Hu, X., Muenke, C., Schulten, K., and Michel, H. 1996. The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum molischianum. Structure* 4:581–597.

Grossman, A. R., Bhaya, D., Apt, K. E., and Kehoe, D. M. 1995. Light-harvesting complexes in oxygenic photosynthesis: Diversity, control, and evolution. *Annu. Rev. Genet.* 29:231–288.

Evolution

Hohmann-Marriott, M. F., and Blankenship, R. E. 2011. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* 62:515–48.

Chen, M., and Zhang, Y. 2008. Tracking the molecular evolution of photosynthesis through characterization of atomic contents of the photosynthetic units. *Photosynth. Res.* 97:255–261.

Iverson, T. M. 2006. Evolution and unique bioenergetic mechanisms in oxygenic photosynthesis. *Curr. Opin. Chem. Biol.* 10:91–100.

Cavalier-Smith, T. 2002. Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr. Biol.* 12:R62–64.

Nelson, N., and Ben-Shem, A. 2005. The structure of photosystem I and evolution of photosynthesis. *BioEssays* 27:914–922.

Green, B. R. 2001. Was "molecular opportunism" a factor in the evolution of different photosynthetic light-harvesting pigment systems? *Proc. Natl. Acad. Sci. U.S.A.* 98:2119–2121.

Dismukes, G. C., Klimov, V. V., Baranov, S. V., Nozlov, Y. N., Das Gupta, J., and Tyryshkin, A. 2001. The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 98:2170–2175.

Moreira, D., Le Guyader, H., and Phillippe, H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405:69–72.

Cavalier-Smith, T. 2000. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* 5:174–182.

Chapter 20

Where to Start

Buchanan, B. B., and Wong, J. H. 2013. A conversation with Andrew Benson: reflections on the discovery of the Calvin–Benson cycle. *Photosynth. Res.* 114:207–214.

Ellis, R. J. 2010. Tackling unintelligent design. *Nature* 463:164–165.

Gutteridge, S., and Pierce, J. 2006. A unified theory for the basis of the limitations of the primary reaction of photosynthetic CO_2 fixation: Was Dr. Pangloss right? *Proc. Natl. Acad. Sci. U.S.A.* 103:7203–7204.

Horecker, B. L. 1976. Unravelling the pentose phosphate pathway. In *Reflections on Biochemistry* (pp. 65–72), edited by A. Kornberg, L. Cornudella, B. L. Horecker, and J. Oro. Pergamon.

Levi, P. 1984. Carbon. In *The Periodic Table.* Random House.

Books and General Reviews

Parry, M. A. J., Andralojc, P. J., Mitchell, R. A. C., Madgwick, P. J., and Keys, A. J. 2003. Manipulation of rubisco: The amount, activity, function and regulation. *J. Exp. Bot.* 54:1321–1333.

Spreitzer, R. J., and Salvucci, M. E. 2002. Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* 53:449–475.

Wood, T. 1985. *The Pentose Phosphate Pathway.* Academic Press.

Buchanan, B. B., Gruissem, W., and Jones, R. L. 2000. *Biochemistry and Molecular Biology of Plants.* American Society of Plant Physiologists.

Enzymes and Reaction Mechanisms

Harrison, D. H., Runquist, J. A., Holub, A., and Miziorko, H. M. 1998. The crystal structure of phosphoribulokinase from *Rhodobacter sphaeroides* reveals a fold similar to that of adenylate kinase. *Biochemistry* 37:5074–5085.

Miziorko, H. M. 2000. Phosphoribulokinase: Current perspectives on the structure/function basis for regulation and catalysis. *Adv. Enzymol. Relat. Areas Mol. Biol.* 74:95–127.

Thorell, S., Gergely, P., Jr., Banki, K., Perl, A., and Schneider, G. 2000. The three-dimensional structure of human transaldolase. *FEBS Lett.* 475:205–208.

Carbon Dioxide Fixation and Rubisco

Satagopan, S., Scott, S. S., Smith, T. G., and Tabita, F. R. 2009. A rubisco mutant that confers growth under a normally "inhibitory" oxygen concentration. *Biochemistry* 48:9076–9083.

Tcherkez, G. G. B., Farquhar, G. D., and Andrews, J. T. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U.S.A.* 103:7246–7251.

Sugawara, H., Yamamoto, H., Shibata, N., Inoue, T., Okada, S., Miyake, C., Yokota, A., and Kai, Y. 1999. Crystal structure of carboxylase reaction-oriented ribulose 1,5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita. J. Biol. Chem.* 274:15655–15661.

Hansen, S., Vollan, V. B., Hough, E., and Andersen, K. 1999. The crystal structure of rubisco from *Alcaligenes eutrophus* reveals a novel central eight-stranded β-barrel formed by β-strands from four subunits. *J. Mol. Biol.* 288:609–621.

Knight, S., Andersson, I., and Branden, C. I. 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution: Subunit interactions and active site. *J. Mol. Biol.* 215:113–160.

Taylor, T. C., and Andersson, I. 1997. The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. *J. Mol. Biol.* 265:432–444.

Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C., and Lorimer, G. H. 1998. Mechanism of rubisco: The carbamate as general base. *Chem. Rev.* 98:549–561.

Buchanan, B. B. 1992. Carbon dioxide assimilation in oxygenic and anoxygenic photosynthesis. *Photosynth. Res.* 33:147–162.

Hatch, M. D. 1987. C$_4$ photosynthesis: A unique blend of modified biochemistry, anatomy, and ultrastructure. *Biochim. Biophys. Acta* 895:81–106.

Regulation

Keown, J. R., Griffin, M. D. W., Mertens, H. D. T., and Pearce, F. G. 2013. Small oligomers of ribulose-bisphosphate carboxylase/oxygenase (rubisco) activase are required for biological activity. *J. Biol. Chem.* 288:20607–20615.

Carmo-Silva, A. E., and Salvucci, M. E. 2013. The regulatory properties of rubisco activase differ among species and affect photosynthetic induction during light transitions. *Plant Physiol.* 161:1645–1655.

Gontero, B., and Maberly, S. C. 2012. An intrinsically disordered protein, CP12: Jack of all trades and master of the Calvin cycle. *Biochem. Soc. Trans.* 40:995–999.

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F.-U., Bracher, A., Hayer-Hartl, M. 2011. Structure of green-type Rubisco activase from tobacco. *Nature Struct. Mol. Biol.* 18:1366–1370.

Lebreton, S., Andreescu, S., Graciet, E., and Gontero, B. 2006. Mapping of the interaction site of CP12 with glyceraldehyde-3-phosphate dehydrogenase from *Chlamydomonas reinhardtii*. Functional consequences for glyceraldehyde-3-phosphate dehydrogenase. *FEBS J.* 273:3358–3369.

Graciet, E., Lebreton, S., and Gontero, B. 2004. The emergence of new regulatory mechanisms in the Benson-Calvin pathway via protein-protein interactions: A glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. *J. Exp. Bot.* 55:1245–1254.

Balmer, Y., Koller, A., del Val, G., Manieri, W., Schürmann, P., and Buchanan, B. B. 2003. Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc. Natl. Acad. Sci. U.S.A.* 100:370–375.

Wedel, N., Soll, J., and Paap, B. K. 1997. CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. *Proc. Natl. Acad. Sci. U.S.A.* 94:10479–10484.

Avilan, L., Lebreton, S., and Gontero, B. 2000. Thioredoxin activation of phosphoribulokinase in a bi-enzyme complex from *Chlamydomonas reinhardtii* chloroplasts. *J. Biol. Chem.* 275:9447–9451.

Irihimovitch, V., and Shapira, M. 2000. Glutathione redox potential modulated by reactive oxygen species regulates translation of rubisco large subunit in the chloroplast. *J. Biol. Chem.* 275:16289–16295.

Glucose 6-phosphate Dehydrogenase

Howes, R. E., Piel, F. B., Patil, A. P., Nyangiri, O. A., Gething, P. W., Dewi, M., Hogg, M. M., Battle, K. E., Padilla, C. D., Baird, et al. 2012. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: A geostatistical model-based map. *PLoS Med.* 9:e1001339.

Wang, X.-T., and Engel, P. C. 2009. Clinical mutants of human glucose 6-phosphate dehydrogenase: Impairment of NADP$^+$ binding affects both folding and stability. *Biochim. Biophys. Acta* 1792:804–809.

Au, S. W., Gover, S., Lam, V. M., and Adams, M. J. 2000. Human glucose-6-phosphate dehydrogenase: The crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Struct. Fold. Des.* 8:293–303.

Salvemini, F., Franze, A., Iervolino, A., Filosa, S., Salzano, S., and Ursini, M. V. 1999. Enhanced glutathione levels and oxidoresistance mediated by increased glucose-6-phosphate dehydrogenase expression. *J. Biol. Chem.* 274:2750–2757.

Tian, W. N., Braunstein, L. D., Apse, K., Pang, J., Rose, M., Tian, X., and Stanton, R. C. 1999. Importance of glucose-6-phosphate dehydrogenase activity in cell death. *Am. J. Physiol.* 276:C1121–C1131.

Tian, W. N., Braunstein, L. D., Pang, J., Stuhlmeier, K. M., Xi, Q. C., Tian, X., and Stanton, R. C. 1998. Importance of glucose-6-phosphate dehydrogenase activity for cell growth. *J. Biol. Chem.* 273:10609–10617.

Ursini, M. V., Parrella, A., Rosa, G., Salzano, S., and Martini, G. 1997. Enhanced expression of glucose-6-phosphate dehydrogenase in human cells sustaining oxidative stress. *Biochem. J.* 323:801–806.

Evolution

Williams, B. P., Aubry S., and Hibberd, J. M. 2012. Molecular evolution of genes recruited into C$_4$ photosynthesis. *Trends Plant Sci.* 4:213–220.

Sage, R. F., Sage, T. L., and Kocacinar, F. 2012. Photorespiration and the evolution of C$_4$ photosynthesis. *Annu. Rev. Plant Biol.* 63:19–47.

Deschamps, P., Haferkamp, I., d'Hulst, C., Neuhaus, H. E., and Ball, S. G. 2008. The relocation of starch metabolism to chloroplasts: When, why and how. *Trends Plant Sci.* 13:574–582.

Coy, J. F., Dubel, S., Kioschis, P., Thomas, K., Micklem, G., Delius, H., and Poustka, A. 1996. Molecular cloning of tissue-specific transcripts of a transketolase-related gene: Implications for the evolution of new vertebrate genes. *Genomics* 32:309–316.

Schenk, G., Layfield, R., Candy, J. M., Duggleby, R. G., and Nixon, P. F. 1997. Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase. *J. Mol. Evol.* 44:552–572.

Notaro, R., Afolayan, A., and Luzzatto, L. 2000. Human mutations in glucose 6-phosphate dehydrogenase reflect evolutionary history. *FASEB J.* 14:485–494.

Wedel, N., and Soll, J. 1998. Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc. Natl. Acad. Sci. U.S.A.* 95:9699–9704.

Martin, W., and Schnarrenberger, C. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: A case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* 32:1–18.

Ku, M. S., Kano-Murakami, Y., and Matsuoka, M. 1996. Evolution and expression of C4 photosynthesis genes. *Plant Physiol.* 111:949–957.

Pereto, J. G., Velasco, A. M., Becerra, A., and Lazcano, A. 1999. Comparative biochemistry of CO$_2$ fixation and the evolution of autotrophy. *Int. Microbiol.* 2:3–10.

Chapter 21

Where to Start

Fisher, E. H. 2013. Cellular regulation by protein phosphorylation. *Biochem. Biophys. Res. Commun.* 430:865–867.

Greenberg, C. C., Jurczak, M. J., Danos, A. M., and Brady, M. J. 2006. Glycogen branches out: New perspectives on the role of glycogen metabolism in the integration of metabolic pathways. *Am. J. Physiol. Endocrinol. Metab.* 291:E1–E8.

Books and General Reviews

Roach, P. J, Depaoli-Roach, A. A., Hurley, T. D., and Tagliabracci, V. S. 2012. Glycogen and its metabolism: Some new developments and old themes. *Biochem. J.* 441:763–787.

Palm, D. C., Rohwer J. M., and Hofmeyr, J.-H. S. 2013. Regulation of glycogen synthase from mammalian skeletal muscle: A unifying view of allosteric and covalent regulation. *FEBS J.* 280:2–27.

Agius, L. 2008. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem. J.* 414:1–18.

Structural Studies

Nadeau, O. W., Lane, L. A., Xu, D., Sage, J., Priddy, T. S., Artigues, A., Villar, M. T., Yang, Q., Robinson, C. V., Zhang, Y. et al. 2012. Structure and location of the regulatory β subunits in the $(\alpha\beta\gamma\delta)_4$ phosphorylase kinase complex. *J. Biol. Chem.* 287:36651–36661.

Horcajada, C., Guinovart, J. J., Fita, I., and Ferrer, J. C. 2006. Crystal structure of an archaeal glycogen synthase: Insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J. Biol. Chem.* 281:2923–2931.

Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. 2004. Crystal structure of glycogen synthase: Homologous enzymes catalyze glycogen synthesis and degradation. *EMBO J.* 23:3196–3205.

Gibbons, B. J., Roach, P. J., and Hurley, T. D. 2002. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* 319:463–477.

Priming of Glycogen Synthesis

Lomako, J., Lomako, W. M., and Whelan, W. J. 2004. Glycogenin: The primer for mammalian and yeast glycogen synthesis. *Biochim. Biophys. Acta* 1673:45–55.

Lin, A., Mu, J., Yang, J., and Roach, P. J. 1999. Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an intersubunit reaction. *Arch. Biochem. Biophys.* 363:163–170.

Roach, P. J., and Skurat, A. V. 1997. Self-glucosylating initiator proteins and their role in glycogen biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* 57:289–316.

Catalytic Mechanisms

Skamnaki, V. T., Owen, D. J., Noble, M. E., Lowe, E. D., Lowe, G., Oikonomakos, N. G., and Johnson, L. N. 1999. Catalytic mechanism of phosphorylase kinase probed by mutational studies. *Biochemistry* 38:14718–14730.

Buchbinder, J. L., and Fletterick, R. J. 1996. Role of the active site gate of glycogen phosphorylase in allosteric inhibition and substrate binding. *J. Biol. Chem.* 271:22305–22309.

Regulation of Glycogen Metabolism

Zhang, T., Wang, S., Lin, Y., Xu, W., Ye, D., Xiong, Y., Zhao, S., and Guan, K.-L. 2012. Acetylation negatively regulates glycogen phosphorylase by recruiting protein phosphatase 1. *Cell Metab.* 15:75–87.

Díaz, A., Martínez-Pons, C., Fita, I., Ferrer, J. C., Guinovart, J. J. 2011. Processivity and subcellular localization of glycogen synthase depend on a non-catalytic high affinity glycogen-binding site. *J. Biol. Chem.* 286:18505–18514.

Bouskila, M., Hunter, R. W., Ibrahim, A. D. F., Delattre, L., Peggie, M., van Diepen, J. A., Voshol, P. J., Jensen, J., Sakamoto, K. 2010. Allosteric regulation of glycogen synthase controls glycogen synthesis in muscle. *Cell Metab.*12:456–466.

Ros, S., García-Rocha, M., Domínguez, J., Ferrer, J. C., Guinovart, J. J. 2009. Control of liver glycogen synthase activity and intracellular distribution by phosphorylation. *J. Biol. Chem.* 284:6370–6378.

Danos, A. M., Osmanovic, S., and Brady, M. J. 2009. Differential regulation of glycogenolysis by mutant protein phosphatase-1 glycogen-targeting subunits. *J. Biol. Chem.* 284:19544–19553.

Pautsch, A., Stadler, N., Wissdorf, O., Langkopf, E., Moreth, W., Streicher, R. 2008. Molecular recognition of the protein phosphatase 1 glycogen targeting subunit by glycogen phosphorylase. *J. Biol. Chem.* 283:8913–8918.

Boulatnikov, I. G., Peters, J. L., Nadeau, O. W., Sage, J. M., Daniels, P. J., Kumar, P., Walsh, D. A., and Carlson, G. M. 2009. Expressed phosphorylase *b* kinase and its $\alpha\gamma\delta$ subcomplex as regulatory models for the rabbit skeletal muscle holoenzyme. *Biochemistry* 48:10183–10191.

Ros, S., García-Rocha, M., Domínguez, J., Ferrer, J. C., and Guinovart, J. J. 2009. Control of liver glycogen synthase activity and intracellular distribution by phosphorylation. *J. Biol. Chem.* 284:6370–6378.

Danos, A. M., Osmanovic, S., and Brady, M. J. 2009. Differential regulation of glycogenolysis by mutant protein phosphatase-1 glycogen-targeting subunits. *J. Biol. Chem.* 284:19544–19553.

Pautsch, A., Stadler, N., Wissdorf, O., Langkopf, E., Moreth, M., and Streicher, R. 2008. Molecular recognition of the protein phosphatase 1 glycogen targeting subunit by glycogen phosphorylase. *J. Biol. Chem.* 283:8913–8918.

Jope, R. S., and Johnson, G. V. W. 2004. The glamour and gloom of glycogen synthase kinase-3. *Trends Biochem. Sci.* 29:95–102.

Doble, B. W., and Woodgett, J. R. 2003. GSK-3: Tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* 116:1175–1186.

Pederson, B. A., Cheng, C., Wilson, W. A., and Roach, P. J. 2000. Regulation of glycogen synthase: Identification of residues involved in regulation by the allosteric ligand glucose-6-P and by phosphorylation. *J. Biol. Chem.* 275:27753–27761.

Melendez, R., Melendez-Hevia, E., and Canela, E. I. 1999. The fractal structure of glycogen: A clever solution to optimize cell metabolism. *Biophys. J.* 77:1327–1332.

Franch, J., Aslesen, R., and Jensen, J. 1999. Regulation of glycogen synthesis in rat skeletal muscle after glycogen-depleting contractile activity: Effects of adrenaline on glycogen synthesis and activation of glycogen synthase and glycogen phosphorylase. *Biochem. J.* 344:231–235.

Aggen, J. B., Nairn, A. C., and Chamberlin, R. 2000. Regulation of protein phosphatase-1. *Chem. Biol.* 7:R13–R23.

Egloff, M. P., Johnson, D. F., Moorhead, G., Cohen, P. T., Cohen, P., and Barford, D. 1997. Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *EMBO J.* 16:1876–1887.

Wu, J., Liu, J., Thompson, I., Oliver, C. J., Shenolikar, S., and Brautigan, D. L. 1998. A conserved domain for glycogen binding in protein phosphatase-1 targeting subunits. *FEBS Lett.* 439:185–191.

Genetic Diseases

Nyhan, W. L., Barshop, B. A., and Ozand, P. T. 2005. *Atlas of Metabolic Diseases*. (2d ed., pp. 373–408). Hodder Arnold.

Chen, Y.-T. 2001. Glycogen storage diseases. In *The Metabolic and Molecular Bases of Inherited Diseases* (8th ed., pp. 1521–1552), edited by C. R. Scriver., W. S. Sly, B. Childs, A. L. Beaudet, D. Valle, K. W. Kinzler, and B. Vogelstein. McGraw-Hill.

Burchell, A., and Waddell, I. D. 1991. The molecular basis of the hepatic microsomal glucose-6-phosphatase system. *Biochim. Biophys. Acta* 1092:129–137.

Lei, K. J., Shelley, L. L., Pan, C. J., Sidbury, J. B., and Chou, J. Y. 1993. Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type Ia. *Science* 262:580–583.

Ross, B. D., Radda, G. K., Gadian, D. G., Rocker, G., Esiri, M., and Falconer-Smith, J. 1981. Examination of a case of suspected McArdle's syndrome by ^{31}P NMR. *New Engl. J. Med.* 304:1338–1342.

Evolution

Holm, L., and Sander, C. 1995. Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. *EMBO J.* 14:1287–1293.

Hudson, J. W., Golding, G. B., and Crerar, M. M. 1993. Evolution of allosteric control in glycogen phosphorylase. *J. Mol. Biol.* 234:700–721.

Rath, V. L., and Fletterick, R. J. 1994. Parallel evolution in two homologues of phosphorylase. *Nat. Struct. Biol.* 1:681–690.

Melendez, R., Melendez-Hevia, E., and Cascante, M. 1997. How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. *J. Mol. Evol.* 45:446–455.

Rath, V. L., Lin, K., Hwang, P. K., and Fletterick, R. J. 1996. The evolution of an allosteric site in phosphorylase. *Structure* 4:463–473.

Chapter 22

Where to Start

Walther, T. C., and Farese Jr., R. V. 2012. Lipid droplets and cellular lipid metabolism. *Annu. Rev. Biochem.* 81:687–714.

Granneman, J. G., and Moore, H.-P. 2008. Location, location: Protein trafficking and lipolysis in adipocytes. *Trends Endocrinol. Metab.* 19:3–9.

Yang, L., Ding, Y., Chen, Y., Zhang, S., Huo, C., Wang, Y., Yu, J., Zhang, P., Na, H., Zhang, H., et al. 2012. The proteomics of lipid droplets: Structure, dynamics, and functions of the organelle conserved from bacteria to humans. *J. Lipid Res.* 53:1245–1253.

Rinaldo, P., Matern, D., and Bennet, M. J. 2002. Fatty acid oxidation disorders. *Annu. Rev. Physiol.* 64:477–502.

Rasmussen, B. B., and Wolfe, R. R. 1999. Regulation of fatty acid oxidation in skeletal muscle. *Annu. Rev. Nutr.* 19:463–484.

Semenkovich, C. F. 1997. Regulation of fatty acid synthase (FAS). *Prog. Lipid Res.* 36:43–53.

Wolf, G. 1996. Nutritional and hormonal regulation of fatty acid synthase. *Nutr. Rev.* 54:122–123.

Books

Lawrence, G. D. 2010. *The Fats of Life: Essential Fatty Acids in Health and Disease.* Rutgers University Press.

Vance, D. E., and Vance, J. E. (Eds.). 2008. *Biochemistry of Lipids, Lipoproteins, and Membranes.* Elsevier.

Stipanuk, M. H. (Ed.). 2006. *Biochemical and Physiological Aspects of Human Nutrition.* Saunders.

Fatty Acid Oxidation

Ross, L. E., Xiao, X., and Lowe, M. E. 2013. Identification of amino acids in human colipase that mediate adsorption to lipid emulsions and mixed micelles. *Biochim. Biophys. Acta* 1831: 1052–1059.

Badin, P. M., Loubière, C., Coonen, M., Louche, K., Tavernier, G., Bourlier, V., Mairal, A., Rustan, A. C., Smith, S. R., Langin, D., et al. 2012. Regulation of skeletal muscle lipolysis and oxidative metabolism by the co-lipase CGI-58. *J. Lipid Res.* 53:839–848.

Yang, X., Lu, X., Lombès, M., Rha, G. B., Chi, Y. I., Guerin, T. M., Smart, E. J., Liu, J. 2010. The G_0/G_1 switch gene 2 regulates adipose lipolysis through association with adipose triglyceride lipase. *Cell Metab.* 11:194–205.

Wang, Y., Mohsen, A.-W., Mihalik, S. J., Goetzman, E. S., Vockley, J. 2010. Evidence for physical association of mitochondrial fatty acid oxidation and oxidative phosphorylation complexes. *J. Biol.Chem.* 285:29834–29841.

Ahmadian, M., Duncan, R. E., and Sul, H. S. 2009. The skinny on fat: Lipolysis and fatty acid utilization in adipocytes. *Trends Endocrinol. Metab.* 20:424–428.

Farese, R. V., Jr., and Walther, T. C. 2009. Lipid droplets finally get a little R-E-S-P-E-C-T. *Cell* 139:855–860.

Goodman, J. L. 2008. The gregarious lipid droplet. *J. Biol. Chem.* 283: 28005–28009.

Saha, P. K., Kojima, H., Marinez-Botas, J., Sunehag, A. L., and Chan, L. 2004. Metabolic adaptations in absence of perilipin. *J. Biol. Chem.* 279:35150–35158.

Barycki, J. J., O'Brien, L. K., Strauss, A. W., and Banaszak, L. J. 2000. Sequestration of the active site by interdomain shifting: Crystallographic and spectroscopic evidence for distinct conformations of L-3-hydroxyacyl-CoA dehydrogenase. *J. Biol. Chem.* 275:27186–27196.

Ramsay, R. R. 2000. The carnitine acyltransferases: Modulators of acyl-CoA-dependent reactions. *Biochem. Soc. Trans.* 28:182–186.

Fatty Acid Synthesis

Sun, T., Hayakawa, K., Bateman, K. S., and Fraser, M. E. 2010. Identification of the citrate-binding site of human ATP-citrate lyase using x-ray crystallography. *J. Biol. Chem.* 285: 27418–27428.

Fan, F., Williams, H. J., Boyer, J. G., Graham, T. L., Zhao, H., Lehr, R., Qi, H., Schwartz, B., Raushel, F. M., and Meek, T. D. 2012. On the catalytic mechanism of human ATP citrate lyase. *Biochemistry* 51:5198−5211.

Chypre, M., Zaidi, N., and Smans, K. 2012. ATP-citrate lyase: A mini-review. *Biochem. Biophys. Res. Commun.* 422:1–4.

Maier, T., Leibundgut, M., and Ban, N. 2008. The crystal structure of a mammalian fatty acid synthase. *Science* 321:1315–1322.

Ming, D., Kong, Y., Wakil, S. J., Brink, J., and Ma, J. 2002. Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc. Natl. Acad. Sci. U.S.A.* 99:7895–7899.

Zhang, Y.-M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. 2001. Identification and analysis of the acyl carrier protein (ACP) docking site on β-ketoacyl-ACP synthase III. *J. Biol. Chem.* 276:8231–8238.

Davies, C., Heath, R. J., White, S. W., and Rock, C. O. 2000. The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli. Struct. Fold. Design* 8:185–195.

Loftus, T. M., Jaworsky, D. E., Frehywot, G. L., Townsend, C. A., Ronnett, G. V., Lane, M. D., and Kuhajda, F. P. 2000. Reduced food intake and body weight in mice treated with fatty acid synthase inhibitors. *Science* 288:2379–2381.

Acetyl CoA Carboxylase

Kim, C.-W., Moon, Y.-A., Park, S. W., Cheng, D., Kwon, H. J., and Horton, J. D. 2010. Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107:9626–9631.

Brownsey, R. W., Boone, A. N., Elliott, J. E., Kulpa, J. E., and Lee, W. M. 2006. Regulation of acetyl-CoA carboxylase. *Biochem. Soc. Trans.* 34:223–227.

Hardie, D. G., Ross, F. A., and Hawley, S. A. 2013. AMP-activated protein kinase: A target for drugs both ancient and modern. *Chem. Biol.* 19:1222–1236.

Munday, M. R. 2002. Regulation of acetyl CoA carboxylase. *Biochem. Soc. Trans.* 30:1059–1064.

Thoden, J. B., Blanchard, C. Z., Holden, H. M., and Waldrop, G. L. 2000. Movement of the biotin carboxylase B-domain as a result of ATP binding. *J. Biol. Chem.* 275:16183–16190.

Eicosanoids

De Caterina, R. 2011. n–3 Fatty acids in cardiovascular disease. *New Engl. J. Med.* 364:2439–2450.

Harizi, H., Corcuff, J.-B., and Gualde, N. 2008. Arachidonic-acid-derived eicosanoids: Roles in biology and immunopathology. *Trends Mol. Med.* 14:461–469.

Nakamura, M. T., and Nara, T. Y. 2004. Structure, function, and dietary regulation of $\Delta 6$, $\Delta 5$, and $\Delta 9$ desaturases. *Annu. Rev. Nutr.* 24:345–376.

Malkowski, M. G., Ginell, S. L., Smith, W. L., and Garavito, R. M. 2000. The productive conformation of arachidonic acid bound to prostaglandin synthase. *Science* 289:1933–1937.

Smith, T., McCracken, J., Shin, Y.-K., and DeWitt, D. 2000. Arachidonic acid and nonsteroidal anti-inflammatory drugs induce conformational changes in the human prostaglandin endoperoxide H2 synthase-2 (cyclooxygenase-2). *J. Biol. Chem.* 275: 40407–40415.

Kalgutkar, A. S., Crews, B. C., Rowlinson, S. W., Garner, C., Seibert, K., and Marnett L. J. 1998. Aspirin-like molecules that covalently inactivate cyclooxygenase-2. *Science* 280:1268–1270.

Lands, W. E. 1991. Biosynthesis of prostaglandins. *Annu. Rev. Nutr.* 11:41–60.

Sigal, E. 1991. The molecular biology of mammalian arachidonic acid metabolism. *Am. J. Physiol.* 260:L13–L28.

Weissmann, G. 1991. Aspirin. *Sci. Am.* 264(1):84–90.

Vane, J. R., Flower, R. J., and Botting, R. M. 1990. History of aspirin and its mechanism of action. *Stroke* (12 suppl.):IV12–IV23.

Genetic Diseases and Cancer

Celestino-Soper, P. B. S., Violante, S., Crawford, E. L., Luo, R., Lionel, A. C., Delaby, E., Cai, G., Sadikovic, B., Lee, K., Lo, C., et al. 2012. A common X-linked inborn error of carnitine biosynthesis may be a risk factor for nondysmorphic autism. *Proc. Natl. Acad. Sci. U.S.A.* 109:7947–7981.

Currie, E., Schulze, A., Zechner, R., Walther, T. C., and Farese, Jr. R. V. 2013. Cellular fatty acid metabolism and cancer. *Cell. Metab.* 18: 153–161.

Lutas, A. and Yellen, G. 2013. The ketogenic diet: Metabolic influences on brain excitability and epilepsy. *Trends Neurosci.* 36:32–40.

Beckers, A., Organe, S., Timmermans, L., Scheys, K., Peeters, A., Brusselmans, K., Verhoeven, G., and Swinnen, J. V. 2007. Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells. *Cancer Res.* 67:8180–8187.

Kuhajda, F. P. 2006. Fatty acid synthase and cancer: New application of an old pathway. *Cancer Res.* 66:5977–5980.

Nyhan, W. L., Barshop, B. A., and Ozand, P. T. 2005. *Atlas of Metabolic Diseases* (2d ed., pp. 339–300). Hodder Arnold.

Roe, C. R., and Coates, P. M. 2001. Mitochondrial fatty acid oxidation disorders. In *The Metabolic and Molecular Bases of Inherited Diseases* (8th ed., pp. 2297–2326), edited by C. R. Scriver., W. S. Sly, B. Childs, A. L. Beaudet, D. Valle, K. W. Kinzler, and B. Vogelstein. McGraw-Hill.

Brivet, M., Boutron, A., Slama, A., Costa, C., Thuillier, L., Demaugre, F., Rabier, D., Saudubray, J. M., and Bonnefont, J. P. 1999. Defects in activation and transport of fatty acids. *J. Inherit. Metab. Dis.* 22:428–441.

Wanders, R. J., van Grunsven, E. G., and Jansen, G. A. 2000. Lipid metabolism in peroxisomes: Enzymology, functions and dysfunctions of the fatty acid α-and β-oxidation systems in humans. *Biochem. Soc. Trans.* 28:141–149.

Wanders, R. J., Vreken, P., den Boer, M. E., Wijburg, F. A., van Gennip, A. H., and Ijist, L. 1999. Disorders of mitochondrial fatty acyl-CoA β-oxidation. *J. Inherit. Metab. Dis.* 22:442–487.

Kerner, J., and Hoppel, C. 1998. Genetic disorders of carnitine metabolism and their nutritional management. *Annu. Rev. Nutr.* 18: 179–206.

Bartlett, K., and Pourfarzam, M. 1998. Recent developments in the detection of inherited disorders of mitochondrial β-oxidation. *Biochem. Soc. Trans.* 26:145–152.

Pollitt, R. J. 1995. Disorders of mitochondrial long-chain fatty acid oxidation. *J. Inherit. Metab. Dis.* 18:473–490.

Chapter 23

Where to Start

Varshavsky, A. 2012. The Ubiquitin System, an Immense Realm. *Annu. Rev. Biochem.* 81:167–76.

Ubiquitin-Mediated Protein Regulation. 2009. *Annu. Rev. Biochem.* 78: A series of reviews on the various roles of ubiquitin.

Torchinsky, Y. M. 1989. Transamination: Its discovery, biological and chemical aspects. *Trends Biochem. Sci.* 12:115–117.

Watford, M. 2003. The urea cycle. *Biochem. Mol. Biol. Ed.* 31: 289–297.

Books

Magnusson, S. 2010. *Life of Pee: The Story of How Urine Got Everywhere*. Aurum.

Bender, D. A. 2012. *Amino Acid Metabolism* (3rd ed.). Wiley-Blackwell.

Lippard, S. J., and Berg, J. M. 1994. *Principles of Bioinorganic Chemistry.* University Science Books.

Walsh, C. 1979. *Enzymatic Reaction Mechanisms.* W. H. Freeman and Company.

Christen, P., and Metzler, D. E. 1985. *Transaminases.* Wiley.

Ubiquitin and the Proteasome

Shemorry, A., Hwang, C.-S., and Varshavsky, A. 2013. Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* 50:540–551.

Liu, C.-W., and Jacobson, A. D. 2013. Functions of the 19S complex in proteasomal degradation. *Trends Biochem. Sci.* 38:103−110.

Ehlinger, A., and Walters, K. J. 2013. Structural insights into proteasome activation by the 19S regulatory particle. *Biochemistry* 52:3618−3628.

Peth, A. Nathan, J. A. and Goldberg, A. L. 2013. The ATP costs and time required to degrade ubiquitinated proteins by the 26 S proteasome. *J. Biol. Chem.* 288:29215–29222.

Tomko, Jr., R. J., and Hochstrasser, M. 2013. Molecular architecture and assembly of the eukaryotic proteasome. *Annu. Rev. Biochem.* 82:415–445.

Komander, D., and Rape, M. 2012. The ubiquitin code. *Annu. Rev. Biochem.* 81:203–229.

Greer, P. L., Hanayama, R., Bloodgood, B. L., Mardinly, A. R., Lipton, D. M., Flavell, S. W., Kim, T.-K., Griffith, E. C., Waldon, Z., Maehr, R., et al. 2010. The Angelman syndrome protein Ube3A regulates synapse development by ubiquitinating Arc. *Cell* 140:704–716.

Peth, A., Besche, H. C., and Goldberg A. L. 2009. Ubiquitinated proteins activate the proteasome by binding to Usp14/Ubp6, which causes 20S gate opening. *Mol. Cell* 36:794–804.

Lin, G., Li, D., Carvalho, L. P. S., Deng, H., Tao, H., Vogt, G., Wu, K., Schneider, J., Chidawanyika, T., Warren, J. D., et al. 2009. Inhibitors selective for mycobacterial versus human proteasomes. *Nature* 461:621–626.

Giasson, B. I., and Lee, V. M.-Y. 2003. Are ubiquitination pathways central to Parkinson's disease? *Cell* 114:1–8.

Pagano, M., and Benmaamar, R. 2003. When protein destruction runs amok, malignancy is on the loose. *Cancer Cell* 4:251–256.

Hochstrasser, M. 2000. Evolution and function of ubiquitin-like protein-conjugation systems. *Nat. Cell Biol.* 2:E153–E157.

Pyridoxal Phosphate-Dependent Enzymes

Eliot, A. C., and Kirsch, J. F. 2004. Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73:383–415.

Mehta, P. K., and Christen, P. 2000. The molecular evolution of pyridoxal-5′-phosphate-dependent enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* 74:129–184.

Schneider, G., Kack, H., and Lindqvist, Y. 2000. The manifold of vitamin B_6 dependent enzymes. *Structure Fold Des.* 8:R1–R6.

Urea Cycle Enzymes

Haeussinger, D., and Sies, H. 2013. Hepatic encephalopathy: Clinical aspects and pathogenetic concept. *Arch. Biochem. Biophys.* 536:97–100.

Li, M., Li, C., Allen, A., Stanley, C. A., and Smith, T. J. 2012. The structure and allosteric regulation of mammalian glutamate dehydrogenase. *Arch. Biochem. Biophys.* 519:69–80.

Nakagawa, T., Lomb, D. J., Haigis, M. C., and Guarente, L. 2009. SIRT5 deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle. *Cell* 137:560–570.

Lawson, F. S., Charlebois, R. L., and Dillon, J. A. 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.* 13:970–977.

McCudden, C. R., and Powers-Lee, S. G. 1996. Required allosteric effector site for *N*-acetylglutamate on carbamoyl-phosphate synthetase I. *J. Biol. Chem.* 271:18285–18294.

Amino Acid Degradation

Li, M., Smith, C. J., Walker, M. T., and Smith, T. J. 2009. Novel inhibitors complexed with glutamate dehydrogenase: Allosteric regulation by control of protein dynamics. *J. Biol. Chem.* 284:22988–23000.

Smith, T. J., and Stanley, C. A. 2008. Untangling the glutamate dehydrogenase allosteric nightmare. *Trends Biochem. Sci.* 33:557–564.

Fusetti, F., Erlandsen, H., Flatmark, T., and Stevens, R. C. 1998. Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria. *J. Biol. Chem.* 273: 16962–16967.

Titus, G. P., Mueller, H. A., Burgner, J., Rodriguez De Cordoba, S., Penalva, M. A., and Timm, D. E. 2000. Crystal structure of human homogentisate dioxygenase. *Nat. Struct. Biol.* 7:542–546.

Erlandsen, H., and Stevens, R. C. 1999. The structural basis of phenylketonuria. *Mol. Genet. Metab.* 68:103–125.

Genetic Diseases

Jayakumar, A. R., Liu, M., Moriyama, M. Ramakrishnan, R., Forbush III, B., Reddy, P. V. V., and Norenberg, M. D. 2008. Na-K-Cl cotransporter-1 in the mechanism of ammonia-induced astrocyte swelling. *J. Biol. Chem.* 283:33874–33882.

Scriver, C. R., and Sly, W. S. (Eds.), Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W., and Vogelstein, B. 2001. *The Metabolic Basis of Inherited Disease* (8th ed.). McGraw-Hill.

Historical Aspects and the Process of Discovery

Cooper, A. J. L., and Meister, A. 1989. An appreciation of Professor Alexander E. Braunstein: The discovery and scope of enzymatic transamination. *Biochimie* 71:387–404.

Garrod, A. E. 1909. *Inborn Errors in Metabolism.* Oxford University Press (reprinted in 1963 with a supplement by H. Harris).

Childs, B. 1970. Sir Archibald Garrod's conception of chemical individuality: A modern appreciation. *New Engl. J. Med.* 282:71–78.

Holmes, F. L. 1980. Hans Krebs and the discovery of the ornithine cycle. *Fed. Proc.* 39:216–225.

Chapter 24

Where to Start

Brewin, N. J. 2013. Legume root nodule symbiosis. *The Biochemist* 35:14–18.

Christen, P., Jaussi, R., Juretic, N., Mehta, P. K., Hale, T. I., and Ziak, M. 1990. Evolutionary and biosynthetic aspects of aspartate aminotransferase isoenzymes and other aminotransferases. *Ann. N. Y. Acad. Sci.* 585:331–338.

Schneider, G., Kack, H., and Lindqvist, Y. 2000. The manifold of vitamin B_6 dependent enzymes. *Structure Fold Des.* 8:R1–R6.

Rhee, S. G., Chock, P. B., and Stadtman, E. R. 1989. Regulation of *Escherichia coli* glutamine synthetase. *Adv. Enzymol. Mol. Biol.* 62:37–92.

Shemin, D. 1989. An illustration of the use of isotopes: The biosynthesis of porphyrins. *Bioessays* 10:30–35.

Books

Wu, G. 2013. *Amino Acids: Biochemistry and Nutrition*. CRC Press.

Bender, D. A. 2012. *Amino Acid Metabolism* (3rd ed.). Wiley-Blackwell.

Jordan, P. M. (Ed.). 1991. *Biosynthesis of Tetrapyrroles.* Elsevier.

Scriver, C. R. (Ed.), Sly, W. S. (Ed.), Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W., and Vogelstein, B. 2001. *The Metabolic Basis of Inherited Disease* (8th ed.). McGraw-Hill.

McMurry, J. E., and Begley, T. P. 2005. *The Organic Chemistry of Biological Pathways.* Roberts and Company.

Blakley, R. L., and Benkovic, S. J. 1989. *Folates and Pterins* (vol. 2). Wiley.

Walsh, C. 1979. *Enzymatic Reaction Mechanisms.* W. H. Freeman and Company.

Nitrogen Fixation

Spatzal, T., Aksoyoglu, M., Zhang, L., Andrade, S. L. A., Schleicher, E., Weber, S., Rees, D. C., Einsle, O. 2011. Evidence for interstitial carbon in nitrogenase FeMo Cofactor. *Science* 334:940.

Lancaster, K. M., Roemelt, M., Ettenhuber, P., Hu, Y., Ribbe, M. W., Neese, F., Bergmann, U., DeBeer, S. 2011. X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* 334:974–977.

Seefeldt, L. C., Hoffman, B. M., and Dean, D. R. 2009. Mechanism of Mo-dependent nitrogenase. *Annu. Rev. Biochem.* 79:701–722.

Halbleib, C. M., and Ludden, P. W. 2000. Regulation of biological nitrogen fixation. *J. Nutr.* 130:1081–1084.

Einsle, O., Tezcan, F. A., Andrade, S. L., Schmid, B., Yoshida, M., Howard, J. B., and Rees, D. C. 2002. Nitrogenase MoFe-protein at 1.16 Å resolution: A central ligand in the FeMo-cofactor. *Science* 297:1696–1700.

Benton, P. M., Laryukhin, M., Mayer, S. M., Hoffman, B. M., Dean, D. R., and Seefeldt, L. C. 2003. Localization of a substrate binding site on the FeMo-cofactor in nitrogenase: Trapping propargyl alcohol with an α-70-substituted MoFe protein. *Biochemistry* 42:9102–9109.

Regulation of Amino Acid Biosynthesis

Li, Y., Zhang, H., Jiang, C., Xu, M., Pang, Y., Feng, J., Xiang, X., Kong, W., Xu, G., Li, Y., et al. 2013. Hyperhomocysteinemia promotes insulin resistance by inducing endoplasmic reticulum stress in adipose tissue. *J. Biol. Chem.* 288:9583–9592.

Eisenberg, D., Gill, H. S., Pfluegl, G. M., and Rotstein, S. H. 2000. Structure-function relationships of glutamine synthetases. *Biochim. Biophys. Acta* 1477:122–145.

Purich, D. L. 1998. Advances in the enzymology of glutamine synthesis. *Adv. Enzymol. Relat. Areas Mol. Biol.* 72:9–42.

Yamashita, M. M., Almassy, R. J., Janson, C. A., Cascio, D., and Eisenberg, D. 1989. Refined atomic model of glutamine synthetase at 3.5 Å resolution. *J. Biol. Chem.* 264:17681–17690.

Schuller, D. J., Grant, G. A., and Banaszak, L. J. 1995. The allosteric ligand site in the $V_{\max}$-type cooperative enzyme phosphoglycerate dehydrogenase. *Nat. Struct. Biol.* 2:69–76.

Rhee, S. G., Park, R., Chock, P. B., and Stadtman, E. R. 1978. Allosteric regulation of monocyclic interconvertible enzyme cascade systems: Use of *Escherichia coli* glutamine synthetase as an experimental model. *Proc. Natl. Acad. Sci. U.S.A.* 75:3138–3142.

Wessel, P. M., Graciet, E., Douce, R., and Dumas, R. 2000. Evidence for two distinct effector-binding sites in threonine deaminase by site-directed mutagenesis, kinetic, and binding experiments. *Biochemistry* 39:15136–15143.

James, C. L., and Viola, R. E. 2002. Production and characterization of bifunctional enzymes: Domain swapping to produce new bifunctional enzymes in the aspartate pathway. *Biochemistry* 41:3720–3725.

Xu, Y., Carr, P. D., Huber, T., Vasudevan, S. G., and Ollis, D. L. 2001. The structure of the PII-ATP complex. *Eur. J. Biochem.* 268:2028–2037.

Krappmann, S., Lipscomb, W. N., and Braus, G. H. 2000. Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 97:13585–13590.

Aromatic Amino Acid Biosynthesis

Brown, K. A., Carpenter, E. P., Watson, K. A., Coggins, J. R., Hawkins, A. R., Koch, M. H., and Svergun, D. I. 2003. Twists and turns: A tale of two shikimate-pathway enzymes. *Biochem. Soc. Trans.* 31:543–547.

Pan, P., Woehl, E., and Dunn, M. F. 1997. Protein architecture, dynamics and allostery in tryptophan synthase channeling. *Trends Biochem. Sci.* 22:22–27.

Sachpatzidis, A., Dealwis, C., Lubetsky, J. B., Liang, P. H., Anderson, K. S., and Lolis, E. 1999. Crystallographic studies of phosphonate-based α-reaction transition-state analogues complexed to tryptophan synthase. *Biochemistry* 38:12665–12674.

Weyand, M., and Schlichting, I. 1999. Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. *Biochemistry* 38:16469–16480.

Crawford, I. P. 1989. Evolution of a biosynthetic pathway: The tryptophan paradigm. *Annu. Rev. Microbiol.* 43:567–600.

Carpenter, E. P., Hawkins, A. R., Frost, J. W., and Brown, K. A. 1998. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. *Nature* 394:299–302.

Schlichting, I., Yang, X. J., Miles, E. W., Kim, A. Y., and Anderson, K. S. 1994. Structural and kinetic analysis of a channel-impaired mutant of tryptophan synthase. *J. Biol. Chem.* 269:26591–26593.

Glutathione

Edwards, R., Dixon, D. P., and Walbot, V. 2000. Plant glutathione *S*-transferases: Enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* 5:193–198.

Lu, S. C. 2000. Regulation of glutathione synthesis. *Curr. Top. Cell Regul.* 36:95–116.

Schulz, J. B., Lindenau, J., Seyfried, J., and Dichgans, J. 2000. Glutathione, oxidative stress and neurodegeneration. *Eur. J. Biochem.* 267:4904–4911.

Lu, S. C. 1999. Regulation of hepatic glutathione synthesis: Current concepts and controversies. *FASEB J.* 13:1169–1183.

Ethylene and Nitric Oxide

Hill, B. G., Dranka, B. P., Baily, S. M., Lancaster, Jr., J. R., and Darley-Usmar, V. M. 2010. What part of NO don't you understand? Some answers to the cardinal questions in nitric oxide biology. *J. Biol. Chem.* 285:19699–19704.

Nisoli, E., Falcone, S., Tonello, C., Cozzi, V., Palomba, L., Fiorani, M., Pisconti, A., Brunelli, S., Cardile, A., Francolini, M., et al. 2004. Mitochondrial biogenesis by NO yields functionally active mitochondria in mammals. *Proc. Natl. Acad. Sci U.S.A.* 101: 16507–16512.

Bretscher, L. E., Li, H., Poulos, T. L. and Griffith, O. W. 2003. Structural characterization and kinetics of nitric oxide synthase inhibition by novel N_5-(iminoalkyl)- and N_5-(iminoalkenyl)-ornithines. *J. Biol. Chem.* 278:46789–46797.

Haendeler, J., Zeiher, A. M., and Dimmeler, S. 1999. Nitric oxide and apoptosis. *Vitam. Horm.* 57:49–77.

Capitani, G., Hohenester, E., Feng, L., Storici, P., Kirsch, J. F., and Jansonius, J. N. 1999. Structure of 1-aminocyclopropane-1-carboxylate synthase, a key enzyme in the biosynthesis of the plant hormone ethylene. *J. Mol. Biol.* 294:745–756.

Hobbs, A. J., Higgs, A., and Moncada, S. 1999. Inhibition of nitric oxide synthase as a potential therapeutic target. *Annu. Rev. Pharmacol. Toxicol.* 39:191–220.

Stuehr, D. J. 1999. Mammalian nitric oxide synthases. *Biochim. Biophys. Acta* 1411:217–230.

Chang, C., and Shockey, J. A. 1999. The ethylene-response pathway: Signal perception to gene regulation. *Curr. Opin. Plant Biol.* 2:352–358.

Theologis, A. 1992. One rotten apple spoils the whole bushel: The role of ethylene in fruit ripening. *Cell* 70:181–184.

Biosynthesis of Porphyrins

Kaasik, K., and Lee, C. C. 2004. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature* 430:467–471.

Leeper, F. J. 1989. The biosynthesis of porphyrins, chlorophylls, and vitamin B_{12}. *Nat. Prod. Rep.* 6:171–199.

Porra, R. J., and Meisch, H.-U. 1984. The biosynthesis of chlorophyll. *Trends Biochem. Sci.* 9:99–104.

Chapter 25

Where to Start

Sutherland, J. D. 2010. Ribonucleotides. *Cold Spring Harb. Perspect. Biol.* 2:a005439.

Ipata, P. L. 2011. Origin, utilization, and recycling of nucleosides in the central nervous system. *Adv. Physiol. Educ.* 35:342–346.

Ordi, J., Alonso, P. L., de Zulueta, J., Esteban, J., Velasco, M., Mas, E., Campo, E., and Fernández, P. L. 2006. The severe gout of Holy Roman Emperor Charles V. *New Eng. J. Med.* 355:516–520.

Kappock, T. J., Ealick, S. E., and Stubbe, J. 2000. Modular evolution of the purine biosynthetic pathway. *Curr. Opin. Chem. Biol.* 4:567–572.

Jordan, A., and Reichard, P. 1998. Ribonucleotide reductases. *Annu. Rev. Biochem.* 67:71–98.

Pyrimidine Biosynthesis

Raushel, F. M., Thoden, J. B., Reinhart, G. D., and Holden, H. M. 1998. Carbamoyl phosphate synthetase: A crooked path from substrates to products. *Curr. Opin. Chem. Biol.* 2:624–632.

Huang, X., Holden, H. M., and Raushel, F. M. 2001. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu. Rev. Biochem.* 70:149–180.

Begley, T. P., Appleby, T. C., and Ealick, S. E. 2000. The structural basis for the remarkable proficiency of orotidine 5′-monophosphate decarboxylase. *Curr. Opin. Struct. Biol.* 10:711–718.

Traut, T. W., and Temple, B. R. 2000. The chemistry of the reaction determines the invariant amino acids during the evolution and divergence of orotidine 5′-monophosphate decarboxylase. *J. Biol. Chem.* 275:28675–28681.

Purine Biosynthesis

Zhao, H., French, J. B., Fang, Y., and Benkovic, S. J. 2013. The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *Chem. Commun.* 49:4444–4452.

Verrier, F., An, S., Ferrie, A. M., Sun, H., Kyoung, M., Deng, H., Fang, Y., and Benkovic, S. J. 2011. GPCRs regulate the assembly of a multienzyme complex for purine biosynthesis. *Nat. Chem. Biol.* 7:909–915.

Mastrangelo, L., Kim, J.-E., Miyanohara, A., Kang, T. H., and Friedmann, T. 2012. Purinergic signaling in human pluripotent stem cells is regulated by the housekeeping gene encoding hypoxanthine guanine phosphoribosyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* 109:3377–3382.

An, S., Kyoung, M., Allen, J. J., Shokat, K. M., and Benkovic, S. J. 2010. Dynamic regulation of a metabolic multi-enzyme complex by protein kinase CK2. *J. Biol. Chem.* 285:11093–11099.

Thoden, J. B., Firestine, S., Nixon, A., Benkovic, S. J., and Holden, H. M. 2000. Molecular structure of *Escherichia coli* PurT-encoded glycinamide ribonucleotide transformylase. *Biochemistry* 39:8791–8802.

McMillan, F. M., Cahoon, M., White, A., Hedstrom, L., Petsko, G. A., and Ringe, D. 2000. Crystal structure at 2.4 Å resolution of *Borrelia burgdorferi* inosine 5′-monophosphate dehydrogenase: Evidence of a substrate-induced hinged-lid motion by loop 6. *Biochemistry* 39:4533–4542.

Levdikov, V. M., Barynin, V. V., Grebenko, A. I., Melik-Adamyan, W. R., Lamzin, V. S., and Wilson, K. S. 1998. The structure of SAICAR synthase: An enzyme in the de novo pathway of purine nucleotide biosynthesis. *Structure* 6:363–376.

Ribonucleotide Reductases

Ahmad, M. F., and Dealwis, C. G. 2013. The structural basis for the allosteric regulation of ribonucleotide reductase. *Prog. Mol. Biol. Transl. Sci.* 117:389–410.

Minnihan, E. C., Nocera, D. G., and Stubbe, J. 2013. Reversible, long-range radical transfer in *E. coli* class Ia ribonucleotide reductase. *Acc. Chem. Res.* 46:2524−2535.

Reichard, P. 2010. Ribonucleotide reductases: Substrate specificity by allostery. *Biochem. Biophys. Res. Commun.* 396:19–23

Avval, F. Z., and Holmgren, A. 2009. Molecular mechanisms of thioredoxin and glutaredoxin as hydrogen donors for mammalian S phase ribonucleotide reductase. *J. Biol. Chem.* 284:8233–8240.

Rofougaran, R., Crona M., Vodnala, M., Sjöberg, B. M., and Hofer, A. 2008. Oligomerization status directs overall activity regulation of the *Escherichia coli* class Ia ribonucleotide reductase. *J. Biol. Chem.* 283:35310–35318.

Nordlund, P., and Reichard, P. 2006. Ribonucleotide reductases. *Annu. Rev. Biochem.* 75:681–706.

Eklund, H., Uhlin, U., Farnegardh, M., Logan, D. T., and Nordlund, P. 2001. Structure and function of the radical enzyme ribonucleotide reductase. *Prog. Biophys. Mol. Biol.* 77:177–268.

Reichard, P. 1997. The evolution of ribonucleotide reduction. *Trends Biochem. Sci.* 22:81–85.

Stubbe, J. 2000. Ribonucleotide reductases: The link between an RNA and a DNA world? *Curr. Opin. Struct. Biol.* 10:731–736.

Logan, D. T., Andersson, J., Sjoberg, B. M., and Nordlund, P. 1999. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* 283:1499–1504.

Tauer, A., and Benner, S. A. 1997. The B_{12}-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila:* An evolutionary solution to the ribonucleotide reductase conundrum. *Proc. Natl. Acad. Sci. U.S.A.* 94:53–58.

Stubbe, J., Nocera, D. G., Yee, C. S. and Chang, M. C. 2003. Radical initiation in the class I ribonucleotide reductase: Long-range proton-coupled electron transfer? *Chem. Rev.* 103:2167–2201.

Stubbe, J., and Riggs-Gelasco, P. 1998. Harnessing free radicals: Formation and function of the tyrosyl radical in ribonucleotide reductase. *Trends Biochem. Sci.* 23:438–443.

Thymidylate Synthase and Dihydrofolate Reductase

Liu, C. T., Hanoian, P., French, J. B., Pringle, T. H., Hammes-Schiffer, S., and Benkovic, S. J. 2013. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc. Natl. Acad. Sci. U.S.A.* 110: 10159–10164.

Abali, E. E., Skacel, N. E., Celikkaya, H., and Hsieh, Y.-C. 2008. Regulation of human dihydrofolate reductase activity and expression. *Vitam. Horm.* 79:267–292.

Schnell, J. R., Dyson, H. J., and Wright, P. E. 2004. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.* 33:119–140.

Li, R., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappilly, F., Turley, S., and Hol, W. G. 2000. Three-dimensional structure of *M. tuberculosis* dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs. *J. Mol. Biol.* 295:307–323.

Liang, P. H., and Anderson, K. S. 1998. Substrate channeling and domain-domain interactions in bifunctional thymidylate synthase-dihydrofolate reductase. *Biochemistry* 37:12195–12205.

Miller, G. P., and Benkovic, S. J. 1998. Stretching exercises: Flexibility in dihydrofolate reductase catalysis. *Chem. Biol.* 5:R105–R113.

Carreras, C. W., and Santi, D. V. 1995. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* 64:721–762.

Defects in Nucleotide Biosynthesis

Grunebaum, E., Cohen, A., and Roifman, C. M. 2013. Recent advances in understanding and managing adenosine deaminase and purine nucleoside phosphorylase deficiencies. *Curr. Opin. Allergy Clin. Immunol.* 13:630–638.

Fu, R., and Jinnah, H. A. 2012. Genotype-phenotype correlations in Lesch-Nyhan Disease: Moving beyond the gene. *J. Biol. Chem.* 287:2997−3008.

Richette, P. and Bardin, T. 2010. Gout. *Lancet* 375:318–328.

Aiuti, A., Cattaneo, F., Galimberti, S., Benninghoff, U., Cassani, B., Callegaro, L., Scaramuzza, S., Andolfi, G., Mirolo, M., Brigida, I., et al. 2009. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *New Engl. J. Med.* 360:447–458.

Jurecka, A. 2009. Inborn errors of purine and pyrimidine metabolism. *J. Inherit. Metab. Dis.* 32:247–263.

Nyhan, W. L., Barshop, B. A., and Ozand, P. T. 2005. *Atlas of Metabolic Diseases.* (2d ed., pp. 429–462). Hodder Arnold.

Scriver, C. R., Sly, W. S., Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W., and Vogelstein, B. (Eds.). 2001. *The Metabolic and Molecular Bases of Inherited Diseases* (8th ed., pp. 2513–2704). McGraw-Hill.

Nyhan, W. L. 1997. The recognition of Lesch-Nyhan syndrome as an inborn error of purine metabolism. *J. Inherited Metab. Dis.* 20:171–178.

Wong, D. F., Harris, J. C., Naidu, S., Yokoi, F., Marenco, S., Dannals, R. F., Ravert, H. T., Yaster, M., Evans, A., Rousset, O., et al. 1996. Dopamine transporters are markedly reduced in Lesch-Nyhan disease in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 93:5539–5543.

Neychev, V. K., and Mitev, V. I. 2004. The biochemical basis of the neurobehavioral abnormalities in the Lesch-Nyhan syndrome: A hypothesis. *Med. Hypotheses* 63:131–134.

Chapter 26

Where to Start

Vickers, K. C., and Remaley, A. T. 2014. HDL and cholesterol: Life after the divorce? *J. Lipid Res.* 55:4–12.

Lambert, G., Sjouke, B., Choque, B., Kastelein, J. J. P., and Hovingh, G. K. 2012. The PCSK9 decade. *J. Lipid Res.* 53:2515–2524.

Brown, M. S., and Goldstein, J. L. 2009. Cholesterol feedback: From Schoenheimer's bottle to Scap's MELADL. *J. Lipid Res.* 50:S15–S27.

Gimpl, G., Burger, K., and Fahrenholz, F. 2002. A closer look at the cholesterol sensor. *Trends Biochem. Sci.* 27:595–599.

Oram, J. F. 2002. Molecular basis of cholesterol homeostasis: Lessons from Tangier disease and ABCA1. *Trends Mol. Med.* 8:168–173.

Endo, A. 1992. The discovery and development of HMG-CoA reductase inhibitors. *J. Lipid Res.* 33:1569–1582.

Books

Vance, J. E., and Vance, D. E. (Eds.). 2008. *Biochemistry of Lipids, Lipoproteins and Membranes.* Elsevier.

Nyhan, W. L., Barshop, B. A., and Al-Aqeel, A. I. 2011. *Atlas of Metabolic Diseases*. (3d ed., pp. 659–780). Hodder Arnold.

Scriver, C. R., Sly, W. S., Childs, B., Beaudet, A. L., Valle, D., Kinzler, K. W., and Vogelstein, B. (Eds.). 2001. *The Metabolic and Molecular Bases of Inherited Diseases* (8th ed., pp. 2707–2960). McGraw-Hill.

Phospholipids and Sphingolipids

Lee, J., Taneva, S. G., Holland, B. W., Tieleman, D. P., and Cornell, R. B. 2014. Structural basis for autoinhibition of CTP: phosphocholine cytidylyltransferase (CCT), the regulatory enzyme in phosphatidylcholine synthesis, by its membrane-binding amphipathic helix. *J. Biol. Chem*. 289:1742–1755.

Tang, W. H. W., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, F., Wu, Y., and Hazen, S. L. 2013. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *New Engl. J. Med*. 368:1575–1584.

Pascual, F., and Carman, G. M. 2013. Phosphatidate phosphatase, a key regulator of lipid homeostasis. *Biochim. Biophys. Acta* 1831:514–522.

Bennett, B. J., de Aguiar Vallim, T. Q., Wang, Z., Shih, D. M., Meng, Y., Gregory, J., Allayee, H., Lee, R., Graham, M., Crooke, R., et al. 2013. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab*. 17:49–60.

Claypool, S. M., and Koehler C. M. 2012. The complexity of cardiolipin in health and disease. *Trends Biochem. Sci*. 37:32–41.

Carman, G. M., and Han, G.-S. 2009. Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. *J. Biol. Chem.* 284:2593–2597.

Bartke, N., and Hannun, Y. A. 2009. Bioactive sphingolipids: Metabolism and function. *J. Lipid Res*. 50:S91–S96.

Lee, J., Johnson, J., Ding, Z., Paetzel, M., and Cornell, R. B. 2009. Crystal structure of a mammalian CTP: Phosphocholine cytidylyltransferase catalytic domain reveals novel active site residues within a highly conserved nucleotidyltransferase fold. *J. Biol. Chem.* 284:33535–33548.

Nye, C. K., Hanson, R. W., and Kalhan, S. C. 2008. Glyceroneogenesis is the dominant pathway for triglyceride glycerol synthesis *in vivo* in the rat. *J. Biol. Chem.* 283:27565–27574.

Biosynthesis of Cholesterol and Steroids

Radhakrishnan, A., Goldstein, J. L., McDonald, J. G., and Brown, M. S. 2008. Switch-like control of SREBP-2 transport triggered by small changes in ER cholesterol: A delicate balance. *Cell Metab.* 8:512–521.

DeBose-Boyd, R. A. 2008. Feedback regulation of cholesterol synthesis: Sterol-accelerated ubiquitination and degradation of HMG CoA reductase. *Cell Res.* 18:609–621.

Hampton, R. Y. 2002. Proteolysis and sterol regulation. *Annu. Rev. Cell Dev. Biol.* 18:345–378.

Kelley, R. I., and Herman, G. E. 2001. Inborn errors of sterol biosynthesis. *Annu. Rev. Genom. Hum. Genet.* 2:299–341.

Istvan, E. S., and Deisenhofer, J. 2001. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* 292:1160–1164.

Lipoproteins and Their Receptors

Gustafsen, C., Kjolby, M., Nyegaard, M., Mattheisen, M., Lundhede, J., Buttenschøn, H., Mors, O., Bentzon, J. F., Madsen, P., Nykjaer, A., et al. 2014. The hypercholesterolemia-risk gene SORT1 facilitates PCSK9 secretion. *Cell Metab*. 19:310–318.

Rye, K-A., Bursill, C. A., Lambert, G., Tabet, F., and Barter, P. J. 2009. The metabolism and anti-atherogenic properties of HDL. *J. Lipid Res*. 50:S195–S200.

Rader, D. J., Alexander, E. T., Weibel, G. L., Billheimer, J., and Rothblat, G. H. 2009. The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. *J. Lipid Res*. 50:S189–S194.

Tall, A. R., Yvan-Charvet, L., Terasaka, N., Pagler, T., and Wang, N. 2008. HDL, ABC transporters, and cholesterol efflux: Implications for the treatment of atherosclerosis. *Cell Metab*. 7:365–375.

Jeon, H., and Blacklow, S. C. 2005. Structure and physiologic function of the low-density lipoprotein receptor. *Annu. Rev. Biochem.* 74:535–562.

Beglova, N., and Blacklow, S. C. 2005. The LDL receptor: How acid pulls the trigger. *Trends Biochem. Sci*. 30:309–316.

Oxygen Activation and P450 Catalysis

Stiles, A. R., McDonald, J. G., Bauman, D. R., and Russell, D. W. 2009. CYP7B1: One cytochrome P450, two human genetic diseases, and multiple physiological functions. *J. Biol. Chem*. 284:28485–28489.

Zhou, S.-F., Liu, J.-P., and Chowbay, B. 2009. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab. Rev*. 4:89–295.

Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., and Jhoti, H. 2004. Crystal structure of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* 305:683–686.

Chapter 27

Where to Start

Cahill, Jr. G. F. 2006. Fuel metabolism in starvation. *Annu. Rev. Nutr.* 26:1–22.

Dunn, R. 2013. Everything you know about calories is wrong. *Sci. Am.* (3) 309:57−59.

Kenny, P. J. 2013. The food addiction. *Sci. Am.* (3) 309:44−49.

Taubes, G. 2013. Which one will make you fat? *Sci. Am.* (3) 309:60−65.

Hardie, D. G. 2012. Organismal carbohydrate and lipid homeostasis. *Cold Spring Harb. Perspect. Biol.* 4:a006031.

Books

Kessler, D. A. 2010. *The End of Overeating: Taking Control of the Insatiable American Appetite.* Rodale.

Wrangham, R. 2009. *Catching Fire: How Cooking Made Us Human*. Basic Books.

Stipanuk, M. H. (Ed.). 2006. *Biochemical, Physiological, & Molecular Aspects of Human Nutrition*. Saunders-Elsevier.

Fell, D. 1997. *Understanding the Control of Metabolism*. Portland Press.

Frayn, K. N. 1996. *Metabolic Regulation: A Human Perspective*. Portland Press.

Poortmans, J. R. (Ed.). 2004. *Principles of Exercise Biochemistry*. Karger.

Harris, R. A., and Crabb, D. W. 2011. Metabolic interrelationships. In *Textbook of Biochemistry with Clinical Correlations* (pp. 839–882), edited by T. M. Devlin. Wiley-Liss.

Caloric Homeostasis

Woods, S. C. 2009. The control of food intake: Behavioral versus molecular perspectives. *Cell Metab.* 9:489–498.

Figlewicz, D. P., and Benoit, S. C. 2009. Insulin, leptin, and food reward: Update 2008. *Am. J. Physiol. Integr. Comp. Physiol.* 296:R9–R19.

Israel, D., and Chua, S. Jr. 2009. Leptin receptor modulation of adiposity and fertility. *Trends Endocrinol. Metab.* 21:10–16.

Meyers, M. G., Cowley, M. A., and Münzberg, H. 2008. Mechanisms of leptin action and leptin resistance. *Annu. Rev. Physiol.* 70:537–556.

Sowers, J. R. 2008. Endocrine functions of adipose tissue: Focus on adiponectin. *Clin. Cornerstone* 9:32–38.

Brehma, B. J., and D'Alessio, D. A. 2008. Benefits of high-protein weight loss diets: Enough evidence for practice? *Curr. Opin. Endocrinol., Diabetes, Obesity* 15:416–421.

Coll, A. P., Farooqi, I. S., and O'Rahillt, S. O. 2007. The hormonal control of food intake. *Cell* 129:251–262.

Muoio, D. M., and Newgard, C. B. 2006. Obesity-related derangements in metabolic regulation. *Annu. Rev. Biochem.* 75:367–401.

Diabetes Mellitus

Lee, J., and Ozcan, U. 2014. Unfolded protein response signaling and metabolic diseases. *J. Biol. Chem.* 289:1203−1211.

Yamauchi, T., and Kadowaki, T. 2013. Adiponectin receptor as a key player in healthy longevity and obesity-related diseases. *Cell Metab.* 17:185−196.

Könner, A. C., and Brüning, J. C. 2012. Selective insulin and leptin resistance in metabolic disorders. *Cell Metab.* 16:144−152.

Zhang, B. B., Zhou, G., and Li, C. 2009. AMPK: An emerging drug target for diabetes and the metabolic syndrome. *Cell Metab.* 9:407–416.

Magkos, F., Yannakoulia, M., Chan, J. L., and Mantzoros, C. S. 2009. Management of the metabolic syndrome and type 2 diabetes through lifestyle modification. *Annu. Rev. Nutr.* 29:8.1–8.34.

Muoio, D. M., and Newgard, C. B. 2008. Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nat. Rev. Mol. Cell. Biol.* 9:193–205.

Leibiger, I. B., Leibiger, B., and Berggren, P.-O. 2008. Insulin signaling in the pancreatic β-cell. *Annu. Rev. Nutr.* 28:233–251.

Doria, A., Patti, M. E., and Kahn, C. R. 2008. The emerging architecture of type 2 diabetes. *Cell Metab.* 8:186–200.

Croker, B. A., Kiu, H., and Nicholson, S. E. 2008. SOCS regulation of the JAK/STAT signalling pathway. *Semin. Cell Dev. Biol.* 19:414–422.

Eizirik, D. L., Cardozo, A. K., and Cnop, M. 2008. The role of endoplasmic reticulum stress in diabetes mellitus. *Endocrinol. Rev.* 29:42–61.

Howard, J. K., and Flier, J. S. 2006. Attenuation of leptin and insulin signaling by SOCS proteins. *Trends Endocrinol. Metab.* 9:365–371.

Lowel, B. B., and Shulman, G. 2005. Mitochondrial dysfunction and type 2 diabetes. *Science* 307:384–387.

Taylor, S. I. 2001. Diabetes mellitus. In *The Metabolic Basis of Inherited Diseases* (8th ed., pp. 1433–1469), edited by C. R. Scriver, W. S. Sly, B. Childs, A. L. Beaudet, D. Valle, K. W. Kinzler, and B. Vogelstein. McGraw-Hill.

Exercise Metabolism

Egan, B., and Zierath, J. R. 2013. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. *Cell Metab.* 17:162−184.

Hood, D. A. 2001. Contractile activity-induced mitochondrial biogenesis in skeletal muscle. *J. Appl. Physiol.* 90:1137–1157.

Shulman, R. G., and Rothman, D. L. 2001. The "glycogen shunt" in exercising muscle: A role for glycogen in muscle energetics and fatigue. *Proc. Natl. Acad. Sci. U.S.A.* 98:457–461.

Gleason, T. 1996. Post-exercise lactate metabolism: A comparative review of sites, pathways, and regulation. *Annu. Rev. Physiol.* 58:556–581.

Holloszy, J. O., and Kohrt, W. M. 1996. Regulation of carbohydrate and fat metabolism during and after exercise. *Annu. Rev. Nutr.* 16:121–138.

Hochachka, P. W., and McClelland, G. B. 1997. Cellular metabolic homeostasis during large-scale change in ATP turnover rates in muscles. *J. Exp. Biol.* 200:381–386.

Horowitz, J. F., and Klein, S. 2000. Lipid metabolism during endurance exercise. *Am. J. Clin. Nutr.* 72:558S–563S.

Wagenmakers, A. J. 1999. Muscle amino acid metabolism at rest and during exercise. *Diabetes Nutr. Metab.* 12:316–322.

Metabolic Adaptations in Starvation

Baverel, G., Ferrier, B., and Martin, M. 1995. Fuel selection by the kidney: Adaptation to starvation. *Proc. Nutr. Soc.* 54:197–212.

MacDonald, I. A., and Webber, J. 1995. Feeding, fasting and starvation: Factors affecting fuel utilization. *Proc. Nutr. Soc.* 54:267–274.

Cahill, G. F., Jr. 1976. Starvation in man. *Clin. Endocrinol. Metab.* 5:397–415.

Sugden, M. C., Holness, M. J., and Palmer, T. N. 1989. Fuel selection and carbon flux during the starved-to-fed transition. *Biochem. J.* 263:313–323.

Ethanol Metabolism

Nagy, L. E. 2004. Molecular aspects of alcohol metabolism: Transcription factors involved in early-induced liver injury. *Annu. Rev. Nutr.* 24:55–78.

Molotkov, A., and Duester, G. 2002. Retinol/ethanol drug interaction during acute alcohol intoxication involves inhibition of retinol metabolism to retinoic acid by alcohol dehydrogenase. *J. Biol. Chem.* 277:22553–22557.

Stewart, S., Jones, D., and Day, C. P. 2001. Alcoholic liver disease: New insights into mechanisms and preventive strategies. *Trends Mol. Med.* 7:408–413.

Lieber, C. S. 2000. Alcohol: Its metabolism and interaction with nutrients. *Annu. Rev. Nutr.* 20:395–430.

Niemela, O. 1999. Aldehyde-protein adducts in the liver as a result of ethanol-induced oxidative stress. *Front. Biosci.* 1:D506–D513.

Riveros-Rosas, H., Julian-Sanchez, A., and Pina, E. 1997. Enzymology of ethanol and acetaldehyde metabolism in mammals. *Arch. Med. Res.* 28:453–471.

Chapter 28

Where to Start

O'Donnell, M., Langston, L. and Stillman, B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* 5:1–13.

Johnson, A., and O'Donnell, M. 2005. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu. Rev. Biochem.* 74:283–315.

Kornberg, A. 1988. DNA replication. *J. Biol. Chem.* 263:1–4.

Wang, J. C. 1982. DNA topoisomerases. *Sci. Am.* 247(1):94–109.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709–715.

Greider, C. W., and Blackburn, E. H. 1996. Telomeres, telomerase, and cancer. *Sci. Am.* 274(2):92–97.

Books

Kornberg, A., and Baker, T. A. 1992. *DNA Replication* (2d ed.). W. H. Freeman and Company.

Bloomfield, V. A., Crothers, D., Tinoco, I., and Hearst, J. 2000. *Nucleic Acids: Structures, Properties and Functions.* University Science Books.

Friedberg, E. C., Walker, G. C., and Siede, W. 1995. *DNA Repair and Mutagenesis.* American Society for Microbiology.

Cozzarelli, N. R., and Wang, J. C. (Eds.). 1990. *DNA Topology and Its Biological Effects.* Cold Spring Harbor Laboratory Press.

DNA Topology and Topoisomerases

Graille, M., Cladiere, L., Durand, D., Lecointe, F., Gadelle, D., Quevillon-Cheruel, S., Vachette, P., Forterre, P., and van Tilbeurgh, H. 2008.

Crystal structure of an intact type II DNA topoisomerase: Insights into DNA transfer mechanisms. *Structure* 16:360–370.

Charvin, G., Strick, T. R., Bensimon, D., and Croquette, V. 2005. Tracking topoisomerase activity at the single-molecule level. *Annu. Rev. Biophys. Biomol. Struct.* 34:201–219.

Sikder, D., Unniraman, S., Bhaduri, T., and Nagaraja, V. 2001. Functional cooperation between topoisomerase I and single strand DNA-binding protein. *J. Mol. Biol.* 306:669–679.

Fortune, J. M., and Osheroff, N. 2000. Topoisomerase II as a target for anticancer drugs: When enzymes stop being nice. *Prog. Nucleic Acid Res. Mol. Biol.* 64:221–253.

Isaacs, R. J., Davies, S. L., Sandri, M. I., Redwood, C., Wells, N. J., and Hickson, I. D. 1998. Physiological regulation of eukaryotic topoisomerase II. *Biochim. Biophys. Acta* 1400:121–137.

Wang, J. C. 1998. Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine. *Q. Rev. Biophys.* 31:107–144.

Baird, C. L., Harkins, T. T., Morris, S. K., and Lindsley, J. E. 1999. Topoisomerase II drives DNA transport by hydrolyzing one ATP. *Proc. Natl. Acad. Sci. U.S.A.* 96:13685–13690.

Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank, K. M., and Cozzarelli, N. R. 1992. Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.* 227:1224–1243.

Fisher, L. M., Austin, C. A., Hopewell, R., Margerrison, M., Oram, M., Patel, S., Wigley, D. B., Davies, G. J., Dodson, E. J., Maxwell, A., et al. 1991. Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* 351:624–629.

Mechanism of Replication

Davey, M. J., and O'Donnell, M. 2000. Mechanisms of DNA replication. *Curr. Opin. Chem. Biol.* 4:581–586.

Keck, J. L., and Berger, J. M. 2000. DNA replication at high resolution. *Chem. Biol.* 7:R63–R71.

Kunkel, T. A., and Bebenek, K. 2000. DNA replication fidelity. *Annu. Rev. Biochem.* 69:497–529.

Waga, S., and Stillman, B. 1998. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* 67:721–751.

Marians, K. J. 1992. Prokaryotic DNA replication. *Annu. Rev. Biochem.* 61:673–719.

DNA Polymerases and Other Enzymes of Replication

Kurth, I., and O'Donnell, M. 2013. New insights into replisome fluidity during chromosome replication. *Trends Biochem. Sci.* 38:195–203.

Nandakumar, J., and Cech, T. R. 2013. Finding the end: Recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell. Biol.* 14:69–82.

Singleton, M. R., Sawaya, M. R., Ellenberger, T., and Wigley, D. B. 2000. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* 101:589–600.

Donmez, I., and Patel, S. S. 2006. Mechanisms of a ring shaped helicase. *Nucleic Acids Res.* 34:4216–4224.

Johnson, D. S., Bai, L., Smith, B. Y., Patel, S. S., and Wang, M. D. 2007. Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* 129:1299–1309.

Lee, S. J., Qimron, U., and Richardson, C. C. 2008. Communication between subunits critical to DNA binding by hexameric helicase of bacteriophage T7. *Proc. Natl. Acad. Sci. U.S.A.* 105:8908–8913.

Toth, E. A., Li, Y., Sawaya, M. R., Cheng, Y., and Ellenberger, T. 2003. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell* 12:1113–1123.

Hubscher, U., Maga, G., and Spadari, S. 2002. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71:133–163.

Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391:251–258.

Arezi, B., and Kuchta, R. D. 2000. Eukaryotic DNA primase. *Trends Biochem. Sci.* 25:572–576.

Jager, J., and Pata, J. D. 1999. Getting a grip: Polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* 9:21–28.

Steitz, T. A. 1999. DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* 274:17395–17398.

Beese, L. S., Derbyshire, V., and Steitz, T. A. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* 260:352–355.

McHenry, C. S. 1991. DNA polymerase III holoenzyme: Components, structure, and mechanism of a true replicative complex. *J. Biol. Chem.* 266:19127–19130.

Kong, X. P., Onrust, R., O'Donnell, M., and Kuriyan, J. 1992. Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69: 425–437.

Polesky, A. H., Steitz, T. A., Grindley, N. D., and Joyce, C. M. 1990. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli. J. Biol. Chem.* 265:14579–14591.

Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. 2000. Crystal structure of NAD^+ dependent DNA ligase: Modular architecture and functional implications. *EMBO J.* 19:1119–1129.

Timson, D. J., and Wigley, D. B. 1999. Functional domains of an NAD^+-dependent DNA ligase. *J. Mol. Biol.* 285:73–83.

Doherty, A. J., and Wigley, D. B. 1999. Functional domains of an ATP-dependent DNA ligase. *J. Mol. Biol.* 285:63–71.

von Hippel, P. H., and Delagoutte, E. 2001. A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell* 104:177–190.

Tye, B. K., and Sawyer, S. 2000. The hexameric eukaryotic MCM helicase: Building symmetry from nonidentical parts. *J. Biol. Chem.* 275:34833–34836.

Marians, K. J. 2000. Crawling and wiggling on DNA: Structural insights to the mechanism of DNA unwinding by helicases. *Struct. Fold. Des.* 5:R227–R235.

Soultanas, P., and Wigley, D. B. 2000. DNA helicases: "Inching forward." *Curr. Opin. Struct. Biol.* 10:124–128.

de Lange, T. 2009. How telomeres solve the end-protection problem. *Science* 326:948–952.

Bachand, F., and Autexier, C. 2001. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. *Mol. Cell Biol.* 21:1888–1897.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. 1999. Mammalian telomeres end in a large duplex loop. *Cell* 97:503–514.

McEachern, M. J., Krauskopf, A., and Blackburn, E. H. 2000. Telomeres and their control. *Annu. Rev. Genet.* 34:331–358.

Mutations and DNA Repair

Yang, W. 2003. Damage repair DNA polymerases Y. *Curr. Opin. Struct. Biol.* 13:23–30.

Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. 2001. Human DNA repair genes. *Science* 291:1284–1289.

Shin, D. S., Chahwan, C., Huffman, J. L., and Tainer, J. A. 2004. Structure and function of the double-strand break repair machinery. *DNA Repair (Amst.)* 3:863–873.

Michelson, R. J., and Weinert, T. 2000. Closing the gaps among a web of DNA repair disorders. *BioEssays* 22:966–969.

Aravind, L., Walker, D. R., and Koonin, E. V. 1999. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* 27:1223–1242.

Mol, C. D., Parikh, S. S., Putnam, C. D., Lo, T. P., and Tainer, J. A. 1999. DNA repair mechanisms for the recognition and removal of damaged DNA bases. *Annu. Rev. Biophys. Biomol. Struct.* 28:101–128.

Parikh, S. S., Mol, C. D., and Tainer, J. A. 1997. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway. *Structure* 5:1543–1550.

Vassylyev, D. G., and Morikawa, K. 1997. DNA-repair enzymes. *Curr. Opin. Struct. Biol.* 7:103–109.

Verdine, G. L., and Bruner, S. D. 1997. How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* 4:329–334.

Bowater, R. P., and Wells, R. D. 2000. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog. Nucleic Acid Res. Mol. Biol.* 66:159–202.

Cummings, C. J., and Zoghbi, H. Y. 2000. Fourteen and counting: Unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* 9: 909–916.

Defective DNA Repair and Cancer

Dever, S. M., White, E. R., Hartman, M. C., and Valerie, K. 2012. BRCA1-directed, enhanced and aberrant homologous recombination: Mechanism and potential treatment strategies. *Cell Cycle* 11:687–94.

Berneburg, M., and Lehmann, A. R. 2001. Xeroderma pigmentosum and related disorders: Defects in DNA repair and transcription. *Adv. Genet.* 43:71–102.

Lambert, M. W., and Lambert, W. C. 1999. DNA repair and chromatin structure in genetic diseases. *Prog. Nucleic Acid Res. Mol. Biol.* 63:257–310.

Buys, C. H. 2000. Telomeres, telomerase, and cancer. *New Engl. J. Med.* 342:1282–1283.

Urquidi, V., Tarin, D., and Goodison, S. 2000. Role of telomerase in cell senescence and oncogenesis. *Annu. Rev. Med.* 51:65–79.

Lynch, H. T., Smyrk, T. C., Watson, P., Lanspa, S. J., Lynch, J. F., Lynch, P. M., Cavalieri, R. J., and Boland, C. R. 1993. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: An updated review. *Gastroenterology* 104:1535–1549.

Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. 1993. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. *Cell* 75:1027–1038.

Ames, B. N., and Gold, L. S. 1991. Endogenous mutagens and the causes of aging and cancer. *Mutat. Res.* 250:3–16.

Ames, B. N. 1979. Identifying environmental chemicals causing mutations and cancer. *Science* 204:587–593.

Recombination and Recombinases

Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., and Wigley, D. B. 2004. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* 432: 187–193.

Spies, M., Bianco, P. R., Dillingham, M. S., Handa, N., Baskin, R. J., and Kowalczykowski, S. C. 2003. A molecular throttle: The recombination hotspot chi controls DNA translocation by the RecBCD helicase. *Cell* 114:647–654.

Kowalczykowski, S. C. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* 25:1562165.

Prevost, C., and Takahashi, M. 2003. Geometry of the DNA strands within the RecA nucleofilament: Role in homologous recombination. *Q. Rev. Biophys.* 36:429–453.

Van Duyne, G. D. 2001. A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* 30:87–104.

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. 2000. Crystal structure of a Flp recombinase-Holliday junction complex: Assembly of an active oligomer by helix swapping. *Mol. Cell* 6:885–897.

Craig, N. L. 1997. Target site selection in transposition. *Annu. Rev. Biochem.* 66:437–474.

Gopaul, D. N., Guo, F., and Van Duyne, G. D. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.* 17:4175–4187.

Gopaul, D. N., and Van Duyne, G. D. 1999. Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* 9:14–20.

Chapter 29

Where to Start

Liu, X., Bushnell, D. A., and Kornberg, R. D. 2013. RNA polymerase II transcription: Structure and mechanism. *Biochim. Biophys. Acta* 1829:2–8.

Kornberg, R. D. 2007. The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* 104:12955–12961.

Woychik, N. A. 1998. Fractions to functions: RNA polymerase II thirty years later. *Cold Spring Harbor Symp. Quant. Biol.* 63:311–317.

Losick, R. 1998. Summary: Three decades after sigma. *Cold Spring Harbor Symp. Quant. Biol.* 63:653–666.

Sharp, P. A. 1994. Split genes and RNA splicing (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* 33:1229–1240.

Cech, T. R. 1990. Nobel lecture: Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. *Biosci. Rep.* 10:239–261.

Villa, T., Pleiss, J. A., and Guthrie, C. 2002. Spliceosomal snRNAs: Mg^{2+} dependent chemistry at the catalytic core? *Cell* 109: 149–152.

Books

Krebs, J. E., and Goldstein, E. S. 2012. *Lewin's Genes XI* (11th ed.). Jones and Bartlett.

Kornberg, A., and Baker, T. A. 1992. *DNA Replication* (2d ed.). W. H. Freeman and Company.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Amon, A., and Scott, M. P. 2012. *Molecular Cell Biology* (7th ed.). W. H. Freeman and Company.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. 2013. *Molecular Biology of the Gene* (7th ed.). Benjamin Cummings.

Gesteland, R. F., Cech, T., and Atkins, J. F. 2006. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA* (3d ed.). Cold Spring Harbor Laboratory Press.

RNA Polymerases

Liu, X., Bushnell, D. A., Wang, D., Calero, G., and Kornberg, R. D. 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327:206–209.

Wang, D., Bushnell, D. A., Huang, X., Westover, K. D., Levitt, M., and Kornberg, R. D. 2009. Structural basis of transcription: Backtracked RNA polymerase II at 3.4 Å resolution. *Science* 324:1203–1206.

Darst, S. A. 2001. Bacterial RNA polymerase. *Curr. Opin. Struct. Biol.* 11:155–162.

Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262:1407–1413.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292:1863–1875.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. 2001. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292:1876–1882.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98:811–824.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104:901–912.

Darst, S. A. 2004. New inhibitors targeting bacterial RNA polymerase. *Trends Biochem. Sci.* 29:159–160.

Cheetham, G. M., and Steitz, T. A. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286:2305–2309.

Ebright, R. H. 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304:687–698.

Paule, M. R., and White, R. J. 2000. Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Res.* 28:1283–1298.

Initiation and Elongation

Murakami, K. S., and Darst, S. A. 2003. Bacterial RNA polymerases: The whole story. *Curr. Opin. Struct. Biol.* 13:31–39.

Buratowski, S. 2000. Snapshots of RNA polymerase II transcription initiation. *Curr. Opin. Cell Biol.* 12:320–325.

Conaway, J. W., and Conaway, R. C. 1999. Transcription elongation and human disease. *Annu. Rev. Biochem.* 68:301–319.

Conaway, J. W., Shilatifard, A., Dvir, A., and Conaway, R. C. 2000. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* 25:375–380.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. 2000. A structural model of transcription elongation. *Science* 289:619–625.

Reines, D., Conaway, R. C., and Conaway, J. W. 1999. Mechanism and regulation of transcriptional elongation by RNA polymerase II. *Curr. Opin. Cell Biol.* 11:342–346.

Promoters, Enhancers, and Transcription Factors

Merika, M., and Thanos, D. 2001. Enhanceosomes. *Curr. Opin. Genet. Dev.* 11:205–208.

Park, J. M., Gim, B. S., Kim, J. M., Yoon, J. H., Kim, H. S., Kang, J. G., and Kim, Y. J. 2001. *Drosophila* mediator complex is broadly utilized by diverse gene-specific transcription factors at different types of core promoters. *Mol. Cell. Biol.* 21:2312–2323.

Smale, S. T., and Kadonaga, J. T. 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72:449–479.

Gourse, R. L., Ross, W., and Gaal, T. 2000. Ups and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* 37:687– 695.

Fiering, S., Whitelaw, E., and Martin, D. I. 2000. To be or not to be active: The stochastic nature of enhancer action. *BioEssays* 22:381–387.

Hampsey, M., and Reinberg, D. 1999. RNA polymerase II as a control panel for multiple coactivator complexes. *Curr. Opin. Genet. Dev.* 9:132–139.

Chen, L. 1999. Combinatorial gene regulation by eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* 9:48–55.

Muller, C. W. 2001. Transcription factors: Global and detailed views. *Curr. Opin. Struct. Biol.* 11:26–32.

Reese, J. C. 2003. Basal transcription factors. *Curr. Opin. Genet. Dev.* 13:114–118.

Kadonaga, J. T. 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116:247–257.

Harrison, S. C. 1991. A structural taxonomy of DNA-binding domains. *Nature* 353:715–719.

Sakurai, H., and Fukasawa, T. 2000. Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 275:37251–37256.

Droge, P., and Muller-Hill, B. 2001. High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells. *Bioessays* 23:179–183.

Smale, S. T., Jain, A., Kaufmann, J., Emami, K. H., Lo, K., and Garraway, I. P. 1998. The initiator element: A paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harbor Symp. Quant. Biol.* 63:21–31.

Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B., 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512–520.

Kim, J. L., Nikolov, D. B., and Burley, S. K., 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520–527.

White, R. J., and Jackson, S. P., 1992. The TATA-binding protein: A central role in transcription by RNA polymerases I, II and III. *Trends Genet.* 8:284–288.

Martinez, E. 2002. Multi-protein complexes in eukaryotic gene transcription. *Plant Mol. Biol.* 50:925–947.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. 2005. A structural perspective of CTD function. *Genes Dev.* 19: 1401–1415.

Palancade, B., and Bensaude, O. 2003. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur. J. Biochem.* 270:3859–3870.

Termination

Burgess, B. R., and Richardson, J. P. 2001. RNA passes through the hole of the protein hexamer in the complex with *Escherichia coli* Rho factor. *J. Biol. Chem.* 276:4182–4189.

Yu, X., Horiguchi, T., Shigesada, K., and Egelman, E. H. 2000. Three-dimensional reconstruction of transcription termination factor rho: Orientation of the N-terminal domain and visualization of an RNA-binding site. *J. Mol. Biol.* 299:1279–1287.

Stitt, B. L. 2001. *Escherichia coli* transcription termination factor Rho binds and hydrolyzes ATP using a single class of three sites. *Biochemistry* 40:2276–2281.

Henkin, T. M. 2000. Transcription termination control in bacteria. *Curr. Opin. Microbiol.* 3:149–153.

Gusarov, I., and Nudler, E. 1999. The mechanism of intrinsic transcription termination. *Mol. Cell* 3:495–504.

Riboswitches

Barrick, J. E., and Breaker, R. R. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 8:R239.

Cheah, M. T., Wachter, A., Sudarsan, N., and Breaker, R. R. 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447:497–500.

Serganov, A., Huang, L., and Patel, D. J. 2009. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 458:233–237.

Noncoding RNA

Cech, T. R. and Steitz, J. A. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157:77–94.

Peculis, B. A. 2002. Ribosome biogenesis: Ribosomal RNA synthesis as a package deal. *Curr. Biol.* 12:R623–R624.

Decatur, W. A., and Fournier, M. J. 2002. rRNA modifications and ribosome function. *Trends Biochem. Sci.* 27:344–351.

Hopper, A. K., and Phizicky, E. M. 2003. tRNA transfers to the limelight. *Genes Dev.* 17:162–180.

Weiner, A. M. 2004. tRNA maturation: RNA polymerization without a nucleic acid template. *Curr. Biol.* 14:R883–R885.

5′-Cap Formation and Polyadenylation

Shatkin, A. J., and Manley, J. L. 2000. The ends of the affair: Capping and polyadenylation. *Nat. Struct. Biol.* 7:838–842.

Bentley, D. L. 2005. Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* 17:251–256.

Aguilera, A. 2005. Cotranscriptional mRNP assembly: From the DNA to the nuclear pore. *Curr. Opin. Cell Biol.* 17:242–250.

Ro-Choi, T. S. 1999. Nuclear snRNA and nuclear function (discovery of 5′ cap structures in RNA). *Crit. Rev. Eukaryotic Gene Expr.* 9:107–158.

Bard, J., Zhelkovsky, A. M., Helmling, S., Earnest, T. N., Moore, C. L., and Bohm, A. 2000. Structure of yeast poly(A) polymerase alone and in complex with 3′-dATP. *Science* 289:1346–1349.

Martin, G., Keller, W., and Doublie, S. 2000. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J.* 19:4193–4203.

Zhao, J., Hyman, L., and Moore, C. 1999. Formation of mRNA 3′ ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* 63:405–445.

Minvielle-Sebastia, L., and Keller, W. 1999. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* 11:352–357.

Small Regulatory RNAs

Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. 2009. Many roads to maturity: MicroRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11:228–234.

Ruvkun, G., Wightman, B., and Ha, I. 2004. The 20 years it took to recognize the importance of tiny RNAs. *Cell* 116:S93–S96.

RNA Editing

Gott, J. M., and Emeson, R. B. 2000. Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* 34:499–531.

Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D., and Maslov, D. A. 2000. Evolution of RNA editing in trypanosome mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 97:6986–6993.

Chester, A., Scott, J., Anant, S., and Navaratnam, N. 2000. RNA editing: Cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim. Biophys. Acta* 1494:1–3.

Maas, S., and Rich, A. 2000. Changing genetic information through RNA editing. *BioEssays* 22:790–802.

Splicing of mRNA Precursors

Caceres, J. F., and Kornblihtt, A. R. 2002. Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Genet*. 18:186–193.

Faustino, N. A., and Cooper, T. A. 2003. Pre-mRNA splicing and human disease. *Genes Dev*. 17:419–437.

Lou, H., and Gagel, R. F. 1998. Alternative RNA processing: Its role in regulating expression of calcitonin/calcitonin gene-related peptide. *J. Endocrinol*. 156:401–405.

Matlin, A. J., Clark, F., and Smith, C. W. 2005. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6:386–398.

McKie, A. B., McHale, J. C., Keen, T. J., Tarttelin, E. E., Goliath, R., van Lith-Verhoeven, J. J., Greenberg, J., Ramesar, R. S., Hoyng, C. B., Cremers, F. P., et al. 2001. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet*. 10:1555–1562.

Nilsen, T. W. 2003. The spliceosome: The most complex macromolecular machine in the cell? *BioEssays* 25:1147–1149.

Rund, D., and Rachmilewitz, E. 2005. β-Thalassemia. *New Engl. J. Med.* 353:1135–1146.

Patel, A. A., and Steitz, J. A. 2003. Splicing double: Insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* 4:960–970.

Sharp, P. A. 2005. The discovery of split genes and RNA splicing. *Trends Biochem. Sci*. 30:279–281.

Valadkhan, S., and Manley, J. L. 2001. Splicing-related catalysis by protein-free snRNAs. *Nature* 413:701–707.

Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419: 182–185.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409:539–542.

Strehler, E. E., and Zacharias, D. A. 2001. Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. *Physiol. Rev.* 81:21–50.

Graveley, B. R. 2001. Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* 17:100–107.

Newman, A. 1998. RNA splicing. *Curr. Biol.* 8:R903–R905.

Reed, R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 12:340–345.

Sleeman, J. E., and Lamond, A. I. 1999. Nuclear organization of pre-mRNA splicing factors. *Curr. Opin. Cell Biol.* 11:372–377.

Black, D. L. 2000. Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. *Cell* 103:367–370.

Collins, C. A., and Guthrie, C. 2000. The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* 7:850–854.

Self-Splicing and RNA Catalysis

Adams, P. L., Stanley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* 430:45–50.

Adams, P. L., Stanley, M. R., Gill, M. L., Kosek, A. B., Wang, J., and Strobel, S. A. 2004. Crystal structure of a group I intron splicing intermediate. *RNA* 10:1867–1887.

Stahley, M. R., and Strobel, S. A. 2005. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 309:1587–1590.

Carola, C., and Eckstein, F. 1999. Nucleic acid enzymes. *Curr. Opin. Chem. Biol.* 3:274–283.

Doherty, E. A., and Doudna, J. A. 2000. Ribozyme structures and mechanisms. *Annu. Rev. Biochem.* 69:597–615.

Fedor, M. J. 2000. Structure and function of the hairpin ribozyme. *J. Mol. Biol.* 297:269–291.

Hanna, R., and Doudna, J. A. 2000. Metal ions in ribozyme folding and catalysis. *Curr. Opin. Chem. Biol.* 4:166–170.

Scott, W. G. 1998. RNA catalysis. *Curr. Opin. Struct. Biol.* 8:720–726.

Chapter 30

Where to Start

Yusupova, G., and Yusupov, M. 2014. High-resolution structure of the eukaryotic 80S ribosome. *Annu. Rev. Biochem.* 83:467–486.

Anger, A. M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson. D. N., and Beckmann, R. 2013. Structures of the human and Drosophila 80S ribosome. *Nature* 497:80–87.

Novoa, E. M., and Ribas de Pouplana, L. 2012. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet*. 28:574−581.

Ibba, M., Curnow, A. W., and Söll, D. 1997. Aminoacyl-tRNA synthesis: Divergent routes to a common goal. *Trends Biochem. Sci.* 22:39–42.

Koonin, E. V., and Novozhilov, A. S. 2009. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* 61:99–111.

Schimmel, P., and Ribas de Pouplana, L. 2000. Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci.* 25:207–209.

Books

Rodnina, M. V., Wintermeyer, W., and Green, R. 2011 (Eds.). *Ribosome Structure, Function and Dynamics*. Springer.

Cold Spring Harbor Symposia on Quantitative Biology. 2001. Volume 66. *The Ribosome*. Cold Spring Harbor Laboratory Press.

Gesteland, R. F., Atkins, J. F., and Cech, T. (Eds.). 2005. *The RNA World* (3d ed.). Cold Spring Harbor Laboratory Press.

Garrett, R., Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B., and Noller, H. F. 2000. *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions.* The American Society for Microbiology.

Aminoacyl-tRNA Synthetases

Kaminska, M., Havrylenko, S., Decottignies, P., Le Maréchal, P., Negrutskii, B., and Mirande, M. 2009. Dynamic organization of aminoacyl-tRNA synthetase complexes in the cytoplasm of human cells. *J. Biol. Chem.* 284:13746–13754.

Park, S. G., Schimmel, P., and Kim, S. 2008. Aminoacyl tRNA synthetases and their connections to disease. *Proc. Natl. Acad. Sci. U.S.A.* 105:11043–11049.

Ibba, M., and Söll, D. 2000. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69:617–650.

Sankaranarayanan, R., Dock-Bregeon, A. C., Rees, B., Bovee, M., Caillet, J., Romby, P., Francklyn, C. S., and Moras, D. 2000. Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat. Struct. Biol.* 7:461–465.

Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. 1999. The structure of threonyl-tRNA synthetase-tRNA$^{\text{Thr}}$ complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* 97:371–381.

Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C., and Moras, D. 2000. Transfer RNA-mediated editing in threonyl-tRNA synthetase: The class II solution to the double discrimination problem. *Cell* 103:877–884.

de Pouplana, L. R., and Schimmel, P. 2000. A view into the origin of life: Aminoacyl-tRNA synthetases. *Cell. Mol. Life Sci.* 57:865–870.

Transfer RNA

Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L., and Söll, D. 2000. The adaptor hypothesis revisited. *Trends Biochem. Sci.* 25:311–316.

Weisblum, B. 1999. Back to Camelot: Defining the specific role of tRNA in protein synthesis. *Trends Biochem. Sci.* 24:247–250.

Ribosomes and Ribosomal RNAs

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., and Ban, N. 2012. Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* 37:189–198.

Jin, H., Kelley, A. C., Loakes, D., and Ramakrishnan, V. 2010. Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc. Natl. Acad. Sci. U.S.A.* 107:8593–8598.

Rodnina, M. V., and Wintermeyer, W. 2009. Recent mechanistic insights into eukaryotic ribosomes. *Curr. Opin. Cell Biol.* 21:435–443.

Dinman, J. D. 2008. The eukaryotic ribosome: Current status and challenges. *J. Biol. Chem.* 284:11761–11765.

Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S. H., Noller, H. F., Bustamante, C., and Tinoco, I., Jr. 2008. Following translation by single ribosomes one codon at a time. *Nature* 452:598–603.

Korostelev, A., and Noller, H. F. 2007. The ribosome in focus: New structures bring insights. *Trends Biochem. Sci.* 32:434–441.

Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., and Baumeister, W. 2009. The native 3D organization of bacterial polysomes. *Cell* 136:261–271.

Initiation Factors

Søgaard, B., Sørensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* 69:101–123.

Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. 2001. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291:498–501.

Guenneugues, M., Caserta, E., Brandi, L., Spurio, R., Meunier, S., Pon, C. L., Boelens, R., and Gualerzi, C. O. 2000. Mapping the fMet-tRNA$_{\text{f}}^{\text{Met}}$ binding site of initiation factor IF2. *EMBO J.* 19:5233–5240.

Meunier, S., Spurio, R., Czisch, M., Wechselberger, R., Guenneugues, M., Gualerzi, C. O., and Boelens, R. 2000. Structure of the fMet-tRNA$_{\text{f}}^{\text{Met}}$-binding domain of *B. stearothermophilus* initiation factor IF2. *EMBO J.* 19:1918–1926.

Elongation Factors

Voorhees R. M., and Ramakrishnan, V. 2013. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* 82:203–236.

Liu, S., Bachran, C., Gupta, P., Miller-Randolph, S., Wang, H., Crown, D., Zhang, Y., Kavaliauskas, D., Nissen, P., and Knudsen, C. R. 2012. The busiest of all ribosomal assistants: Elongation factor Tu. *Biochemistry* 51:2642−2651.

Schuette, J.-C., Murphy, F. V., Kelley, A. C., Weir, J. R., Giesebrecht, J., Connell, S. R., Loerke, J., Mielke, T., Zhang, W., Penczek, P. A., et al. 2009. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J.* 28:755–765.

Stark, H., Rodnina, M. V., Wieden, H. J., van Heel, M., and Wintermeyer, W. 2000. Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* 100:301–309.

Baensch, M., Frank, R., and Kohl, J. 1998. Conservation of the aminoterminal epitope of elongation factor Tu in Eubacteria and Archaea. *Microbiology* 144:2241–2246.

Krasny, L., Mesters, J. R., Tieleman, L. N., Kraal, B., Fucik, V., Hilgenfeld, R., and Jonak, J. 1998. Structure and expression of elongation factor Tu from *Bacillus stearothermophilus. J. Mol. Biol.* 283:371–381.

Pape, T., Wintermeyer, W., and Rodnina, M. V. 1998. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J.* 17:7490–7497.

Piepenburg, O., Pape, T., Pleiss, J. A., Wintermeyer, W., Uhlenbeck, O. C., and Rodnina, M. V. 2000. Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome. *Biochemistry* 39:1734–1738.

Peptide-Bond Formation and Translocation

Rodnina, M. V. 2013. The ribosome as a versatile catalyst: Reactions at the peptidyl transferase center. *Curr. Opin. Struct. Biol.* 23:595–602.

Uemura, S., Aitken, C. E., Korlach, J., Flusberg, B. A., Turner, S. W., and Puglisi, J. D. 2010. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* 464:1012–1018.

Beringer, M., and. Rodnina, M. V. 2007. The ribosomal peptidyl transferase. *Mol. Cell* 26:311–321.

Yarus, M., and Welch, M. 2000. Peptidyl transferase: Ancient and exiguous. *Chem. Biol.* 7:R187–R190.

Vladimirov, S. N., Druzina, Z., Wang, R., and Cooperman, B. S. 2000. Identification of 50S components neighboring 23S rRNA nucleotides A2448 and U2604 within the peptidyl transferase center of *Escherichia coli* ribosomes. *Biochemistry* 39:183–193.

Frank, J., and Agrawal, R. K. 2000. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406:318–322.

Termination

Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R. M., Petry, S., Kelley, A. C., and Ramakrishnan, V. 2008. Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* 322:953–956.

Trobro, S., and Åqvist, S. 2007. A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol. Cell* 27:758–766.

Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W. G., and Noller, H. F. 2008. Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl. Acad. Sci. U.S.A.* 105: 19684–19689.

Wilson, D. N., Schluenzen, F., Harms, J. M., Yoshida, T., Ohkubo, T., Albrecht, A., Buerger, J., Kobayashi, Y., and Fucini, P. 2005. X-ray crystallography study on ribosome recycling: The mechanism of binding and action of RRF on the 50S ribosomal subunit. *EMBO J.* 24:251–260.

Kisselev, L. L., and Buckingham, R. H. 2000. Translational termination comes of age. *Trends Biochem. Sci.* 25:561–566.

Fidelity and Proofreading

Zaher, H. S., and Green, R. 2009. Quality control by the ribosome following peptide bond formation. *Nature* 457:161–166.

Zaher, H. S., and Green, R. 2009. Fidelity at the molecular level: Lessons from protein synthesis. *Cell* 136:746–762.

Ogle, J. M., and Ramakrishnan, V. 2005. Structural insights into translational fidelity. *Annu. Rev. Biochem.* 74:129–177.

Eukaryotic Protein Synthesis

Hinnebusch, A. G. 2014. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* 83:779–812.

Wein, A. N., Singh, R., Fattah, R., and Leppla, S. H. 2012. Diphthamide modification on eukaryotic elongation factor 2 is needed to assure fidelity of mRNA translation and mouse development. *Proc. Natl. Acad. Sci. U.S.A.* 109:13817–13822.

Rhoads, R. E. 2009. eIF4E: New family members, new binding partners, new roles. *J. Biol. Chem.* 284:16711–16715.

Marintchev, A., Edmonds, K. A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N., and Wagner, G. 2009. Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* 136:447–460.

Fitzgerald, K. D., and Semler, B. L. 2009. Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochim. Biophys. Acta* 1789:518–528.

Mitchell, S. F., and Lorsch, J. R. 2008. Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. *J. Biol. Chem.* 283:27345–27349.

Amrani, A., Ghosh, S., Mangus, D. A., and Jacobson, A. 2008. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* 453:1276–1280.

Sachs, A. B., and Varani, G. 2000. Eukaryotic translation initiation: There are (at least) two sides to every story. *Nat. Struct. Biol.* 7:356–361.

Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234:187–208.

Bushell, M., Wood, W., Clemens, M. J., and Morley, S. J. 2000. Changes in integrity and association of eukaryotic protein synthesis initiation factors during apoptosis. *Eur. J. Biochem.* 267:1083–1091.

Das, S., Ghosh, R., and Maitra, U. 2001. Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J. Biol. Chem.* 276:6720–6726.

Lee, J. H., Choi, S. K., Roll-Mecak, A., Burley, S. K., and Dever, T. E. 1999. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. *Proc. Natl. Acad. Sci. U.S.A.* 96:4342–4347.

Pestova, T. V., and Hellen, C. U. 2000. The structure and function of initiation factors in eukaryotic protein synthesis. *Cell. Mol. Life Sci.* 57:651–674.

Antibiotics and Toxins

Belova, L., Tenson, T., Xiong, L., McNicholas, P. M., and Mankin, A. S. 2001. A novel site of antibiotic action in the ribosome: Interaction of evernimicin with the large ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.* 98:3726–3731.

Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. 2000. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* 103:1143–1154.

Porse, B. T., and Garrett, R. A. 1999. Ribosomal mechanics, antibiotics, and GTP hydrolysis. *Cell* 97:423–426.

Lord, M. J., Jolliffe, N. A., Marsden, C. J., Pateman, C. S., Smith, D. S., Spooner, R. A., Watson, P. D., and Roberts, L. M. 2003. Ricin: Mechanisms of toxicity. *Toxicol. Rev.* 22:53–64.

Protein Transport Across Membranes

Akopian, D., Shen, K., Zhang, X., and Shan, S. 2013. Signal recognition particle: An essential protein-targeting machine. *Annu. Rev. Biochem.* 82:693–721.

Nyathi, Y., Wilkinson, B. M., and Pool, M. R. 2013. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim. Biophys. Acta* 1833:2392–2402.

Janda, C. Y., Li, J., Oubridge, C., Hernández, H., Robinson, C. V., and Nagai, K. 2010. Recognition of a signal peptide by the signal recognition particle. *Nature* 465:507–510.

Cross, B. C. S., Sinning, I., Luirink, J., and High, S. 2009. Delivering proteins for export from the cytosol. *Nat. Rev. Mol. Cell. Biol.* 10:255–264.

Shan, S., Schmid, S. L., and Zhang, X. 2009. Signal recognition particle (SRP) and SRP receptor: A new paradigm for multistate regulatory GTPases. *Biochemistry* 48:6696–6704.

Johnson, A. E. 2009. The structural and functional coupling of two molecular machines, the ribosome and the translocon. *J. Cell Biol.* 185:765–767.

Pool, R. P. 2009. A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J. Cell Biol.* 185:889–902.

Egea, P. F., Stroud, R. M., and Walter, P. 2005. Targeting proteins to membranes: Structure of the signal recognition particle. *Curr. Opin. Struct. Biol.* 15:213–220.

Halic, M., and Beckmann, R. 2005. The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.*15:116–125.

Doudna, J. A., and Batey, R. T. 2004. Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* 73:539–557.

Schnell, D. J., and Hebert, D. N. 2003. Protein translocons: Multifunctional mediators of protein translocation across membranes. *Cell* 112:491–505.

Chapter 31

Where to Start

Ptashne, M. 2014. The chemistry of regulation of genes and other things. *J. Biol. Chem.* 289:5417–5435.

Pabo, C. O., and Sauer, R. T. 1984. Protein–DNA recognition. *Annu. Rev. Biochem.* 53:293–321.

Ptashne, M., Johnson, A. D., and Pabo, C. O. 1982. A genetic switch in a bacterial virus. *Sci. Am.* 247:128–140.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M., and Sauer, R. T. 1980. How the lambda repressor and Cro work. *Cell* 19:1–11.

Books

Ptashne, M. 2004. *A Genetic Switch: Phage 3 Revisited* (3d ed.). Cold Spring Harbor Laboratory Press.

McKnight, S. L., and Yamamoto, K. R. (Eds.). 1992. *Transcriptional Regulation* (vols. 1 and 2). Cold Spring Harbor Laboratory Press.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Pleogh, H., Amon, A., and Scott, M. P. 2012. *Molecular Cell Biology* (7th ed.). W. H. Freeman and Company.

DNA-Binding Proteins

Balaeff, A., Mahadevan, L., and Schulten, K. 2004. Structural basis for cooperative DNA binding by CAP and *lac* repressor. *Structure* 12:123–132.

Bell, C. E., and Lewis, M. 2001. The Lac repressor: A second generation of structural and functional studies. *Curr. Opin. Struct. Biol.* 11:19–25.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–1254.

Niu, W., Kim, Y., Tau, G., Heyduk, T., and Ebright, R. H. 1996. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell* 87:1123–1134.

Schultz, S. C., Shields, G. C., and Steitz, T. A. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees. *Science* 253:1001–1007.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E., and Berman, H. M. 1996. Structure of the CAP-DNA complex at 2.5 Å resolution: A complete picture of the protein–DNA interface. *J. Mol. Biol.* 260:395–408.

Busby, S., and Ebright, R. H. 1999. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* 293:199–213.

Somers, W. S., and Phillips, S. E. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature* 359:387–393.

Gene-Regulatory Circuits

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. 1981. Lambda repressor and Cro: Components of an efficient molecular switch. *Nature* 294:217–223.

Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M. 2008. Crystal structure of the lambda repressor and a model for pairwise cooperative operator binding. *Nature* 452:1022–1025.

Arkin, A., Ross, J., and McAdams, H. H. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149:1633–1648.

Post-transcriptional Regulation

Kolter, R., and Yanofsky, C. 1982. Attenuation in amino acid biosynthetic operons. *Annu. Rev. Genet.* 16:113–134.

Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289:751–758.

Miller, M. B., and Bassler, B. L. 2001. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* 55:165–199.

Zhang, R. G., Pappas, T., Brace, J. L., Miller, P. C., Oulmassov, T., Molyneaux, J. M., Anderson, J. C., Bashkin, J. K., Winans, S. C., and Joachimiak, A. 2002. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature* 417:971–974.

Soberon-Chavez, G., Aguirre-Ramirez, M., and Ordonez, L. 2005. Is *Pseudomonas aeruginosa* only "sensing quorum"? *Crit. Rev. Microbiol.* 31:171–182.

Historical Aspects

Lewis, M. 2005. The lac repressor. *C. R. Biol.* 328:521–548.

Jacob, F., and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318–356.

Ptashne, M., and Gilbert, W. 1970. Genetic repressors. *Sci. Am.* 222(6):36–44.

Lwoff, A., and Ullmann, A. (Eds.). 1979. *Origins of Molecular Biology: A Tribute to Jacques Monod*. Academic Press.

Judson, H. 1996. *The Eighth Day of Creation: Makers of the Revolution in Biology*. Cold Spring Harbor Laboratory Press.

Chapter 32

Where to Start

Liu, X., Bushnell, D. A., and Kornberg, R. D. 2013. RNA polymerase II transcription: Structure and mechanism. *Biochim. Biophys. Acta* 1829:2–8.

Kornberg, R. D. 2007. The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* 104:12955–12961.

Pabo, C. O., and Sauer, R. T. 1984. Protein–DNA recognition. *Annu. Rev. Biochem.* 53:293–321.

Struhl, K. 1989. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* 14:137–140.

Struhl, K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98:1–4.

Korzus, E., Torchia, J., Rose, D. W., Xu, L., Kurokawa, R., McInerney, E. M., Mullen, T. M., Glass, C. K., and Rosenfeld, M. G. 1998. Transcription factor-specific requirements for coactivators and their acetyltransferase functions. *Science* 279:703–707.

Aalfs, J. D., and Kingston, R. E. 2000. What does "chromatin remodeling" mean? *Trends Biochem. Sci.* 25:548–555.

Books

McKnight, S. L., and Yamamoto, K. R. (Eds.). 1992. *Transcriptional Regulation* (vols. 1 and 2). Cold Spring Harbor Laboratory Press.

Latchman, D. S. 2004. *Eukaryotic Transcription Factors* (4th ed.). Academic Press.

Wolffe, A. 1992. *Chromatin Structure and Function*. Academic Press.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Pleogh, H., Amon, A., and Scott, M. P. 2012. *Molecular Cell Biology* (7th ed.). W. H. Freeman and Company.

Chromatin and Chromatin Remodeling

Sadeh, R. and Allis, C. D. 2011. Genome-wide "re"-modeling of nucleosome positions. *Cell* 147:263–266.

Lorch, Y., Maier-Davis, B., and Kornberg, R. D. 2010. Mechanism of chromatin remodeling. *Proc. Natl. Acad. Sci. U.S.A.* 107:3458–3462.

Tang, L., Nogales, E., and Ciferri, C. 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* 102:122–128.

Jenuwein, T., and Allis, C. D. 2001. Translating the histone code. *Science* 293:1074–1080.

Jiang, C., and Pugh, B. F. 2009. Nucleosome positioning and gene regulation: Advances through genomics. *Nat. Rev. Genet.* 10:161–172.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.

Weintraub, H., Larsen, A., and Groudine, M. 1981. β-Globin-gene switching during the development of chicken embryos: Expression and chromosome structure. *Cell* 24:333–344.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA-binding proteins. *Science* 290:2306–2309.

Goodrich, J. A., and Tjian, R. 1994. TBP-TAF complexes: Selectivity factors for eukaryotic transcription. *Curr. Opin. Cell. Biol.* 6:403–409.

Bird, A. P., and Wolffe, A. P. 1999. Methylation-induced repression: Belts, braces, and chromatin. *Cell* 99:451–454.

Cairns, B. R. 1998. Chromatin remodeling machines: Similar motors, ulterior motives. *Trends Biochem. Sci.* 23:20–25.

Albright, S. R., and Tjian, R. 2000. TAFs revisited: More data reveal new twists and confirm old ideas. *Gene* 242:1–13.

Urnov, F. D., and Wolffe, A. P. 2001. Chromatin remodeling and transcriptional activation: The cast (in order of appearance). *Oncogene* 20:2991–3006.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260.

Arents, G., and Moudrianakis, E. N. 1995. The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. U.S.A.* 92:11170–11174.

Baxevanis, A. D., Arents, G., Moudrianakis, E. N., and Landsman, D. 1995. A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res.* 23:2685–2691.

Transcription Factors

Green, M. R. 2005. Eukaryotic transcription activation: Right on target. *Mol. Cell* 18:399–402.

Kornberg, R. D. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* 30:235–239.

Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., and Marmorstein, R. 1999. Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J.* 18:3521–3532.

Deckert, J., and Struhl, K. 2001. Histone acetylation at promoters is differentially affected by specific activators and repressors. *Mol. Cell. Biol.* 21:2726–2735.

Dutnall, R. N., Tafrov, S. T., Sternglanz, R., and Ramakrishnan, V. 1998. Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related *N*-acetyltransferase superfamily. *Cell* 94:427–438.

Finnin, M. S., Donigian, J. R., Cohen, A., Richon, V. M., Rifkind, R. A., Marks, P. A., Breslow, R., and Pavletich, N. P. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* 401:188–193.

Finnin, M. S., Donigian, J. R., and Pavletich, N. P. 2001. Structure of the histone deacetylase SIR2. *Nat. Struct. Biol.* 8:621–625.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science* 288:1422–1425.

Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D., and Marmorstein, R. 1999. Structure of *Tetrahymena* GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* 401:93–98.

Induced Pluripotent Stem Cells

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872.

Takahashi, K., and Yamanaka, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676.

Park, I. H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K., and Daley, G. Q. 2008. Disease-specific induced pluripotent stem cells. *Cell* 134:877–886.

Yamanaka, S. 2009. A fresh look at iPS cells. *Cell* 137:13–17.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I., and Thomson, J. A. 2009. Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 324:797–801.

Nuclear Hormone Receptors

Downes, M., Verdecia, M. A., Roecker, A. J., Hughes, R., Hogenesch, J. B., Kast-Woelbern, H. R., Bowman, M. E., Ferrer, J. L., Anisfeld, A. M., Edwards, et al. 2003. A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell* 11:1079–1092.

Evans, R. M. 2005. The nuclear receptor superfamily: A Rosetta stone for physiology. *Mol. Endocrinol.* 19:1429–1438.

Xu, W., Cho, H., Kadam, S., Banayo, E. M., Anderson, S., Yates, J. R., III, Emerson, B. M., and Evans, R. M. 2004. A methylation-mediator complex in hormone signaling. *Genes Dev.* 18:144–156.

Evans, R. M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* 240:889–895.

Yamamoto, K. R. 1985. Steroid receptor regulated transcription of specific genes and gene networks. *Annu. Rev. Genet.* 19:209–252.

Tanenbaum, D. M., Wang, Y., Williams, S. P., and Sigler, P. B. 1998. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Natl. Acad. Sci. U.S.A.* 95:5998–6003.

Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* 75:567–578.

Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., and Greene, G. L. 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95:927–937.

Collingwood, T. N., Urnov, F. D., and Wolffe, A. P. 1999. Nuclear receptors: Coactivators, corepressors and chromatin remodeling in the control of transcription. *J. Mol. Endocrinol.* 23:255–275.

Posttranscriptional Regulation

Rouault, T. A., Stout, C. D., Kaptain, S., Harford, J. B., and Klausner, R. D. 1991. Structural relationship between an iron-regulated RNA-binding protein (IRE-BP) and aconitase: Functional implications. *Cell* 64:881–883.

Klausner, R. D., Rouault, T. A., and Harford, J. B. 1993. Regulating the fate of mRNA: The control of cellular iron metabolism. *Cell* 72:19–28.

Gruer, M. J., Artymiuk, P. J., and Guest, J. R. 1997. The aconitase family: Three structural variations on a common theme. *Trends Biochem. Sci.* 22:3–6.

Theil, E. C. 1994. Iron regulatory elements (IREs): A family of mRNA non-coding sequences. *Biochem. J.* 304:1–11.

MicroRNAs

Ruvkun, G. 2008. The perfect storm of tiny RNAs. *Nat. Med.* 14:1041–1045.

Sethupathy, P., and Collins, F. S. 2008. MicroRNA target site polymorphisms and human disease. *Trends Genet.* 24:489–497.

Adams, B. D., Cowee, D. M., and White, B. A. 2009. The role of miR-206 in the epidermal growth factor (EGF) induced repression of estrogen receptor-α (ERα) signaling and a luminal phenotype in MCF-7 breast cancer cells. *Mol. Endocrinol.* 23:1215–1230.

Jegga, A. G., Chen, J., Gowrisankar, S., Deshmukh, M. A., Gudivada, R., Kong, S., Kaimal, V., and Aronow, B. J. 2007. GenomeTrafac: A whole genome resource for the detection of transcription factor binding site clusters associated with conventional and microRNA encoding genes conserved between mouse and human gene orthologs. *Nucleic Acids Res.* 35:D116–D121.

Chapter 33

Where to Start

Axel, R. 1995. The molecular logic of smell. *Sci. Am.* 273(4):154–159.

Dulac, C. 2000. The physiology of taste, vintage 2000. *Cell* 100:607–610.

Yarmolinsky, D. A., Zuker, C. S., and Ryba, N. J. (2009) Common sense about taste: From mammals to insects. *Cell* 139:234–244.

Stryer, L. 1996. Vision: From photon to perception. *Proc. Natl. Acad. Sci. U.S.A.* 93:557–559.

Hudspeth, A. J. 1989. How the ear's works work. *Nature* 341:397–404.

Olfaction

Buck, L., and Axel, R. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 65:175–187.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H., and Mainland, J. D. 2009. Odor coding by a mammalian receptor repertoire. *Sci. Signal.* 2:ra9.

Malnic, B., Hirono, J., Sato, T., and Buck, L. B. 1999. Combinatorial receptor codes for odors. *Cell* 96:713–723.

Zou, D. J., Chesler, A., and Firestein, S. 2009. How the olfactory bulb got its glomeruli: A just so story? *Nat. Rev. Neurosci.* 10:611–618.

De la Cruz, O., Blekhman, R., Zhang, X., Nicolae, D., Firestein, S., and Gilad, Y. 2009. A signature of evolutionary constraint on a subset of ectopically expressed olfactory receptor genes. *Mol. Biol. Evol.* 26:491–494.

Mombaerts, P., Wang, F., Dulac, C., Chao, S. K., Nemes, A., Mendelsohn, M., Edmondson, J., and Axel, R. 1996. Visualizing an olfactory sensory map. *Cell* 87:675–686.

Buck, L. 2005. Unraveling the sense of smell (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* 44:6128–6140.

Belluscio, L., Gold, G. H., Nemes, A., and Axel, R. 1998. Mice deficient in $G_{(\text{olf})}$ are anosmic. *Neuron* 20:69–81.

Vosshall, L. B., Wong, A. M., and Axel, R. 2000. An olfactory sensory map in the fly brain. *Cell* 102:147–159.

Lewcock, J. W., and Reed, R. R. 2003. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl. Acad. Sci. U.S.A.*101:1069–1074.

Reed, R. R. 2004. After the holy grail: Establishing a molecular mechanism for mammalian olfaction. *Cell* 116:329–336.

Taste

Chandrashekar, J., Yarmolinsky, D., von Buchholtz, L., Oka, Y., Sly, W., Ryba, N. J., and Zuker, C. S. 2009. The taste of carbonation. *Science* 326:443–445.

Chandrashekar, J., Hoon, M. A., Ryba, N. J., and Zuker, C. S. 2006. The receptors and cells for mammalian taste. *Nature* 444:288–294.

Huang, A. L., Chen, X., Hoon, M. A., Chandrashekar, J., Guo, W., Tranker, D., Ryba, N. J., and Zuker, C. S. 2006. The cells and logic for mammalian sour taste detection. *Nature* 442:934–938.

Zhao, G. Q., Zhang, Y., Hoon, M. A., Chandrashekar, J., Erlenbach, I., Ryba, N. J. P., and Zuker, C. S. 2003. The receptors for mammalian sweet and umami taste. *Cell* 115:255–266.

Herness, M. S., and Gilbertson, T. A. 1999. Cellular mechanisms of taste transduction. *Annu. Rev. Physiol.* 61:873–900.

Adler, E., Hoon, M. A., Mueller, K. L., Chandrashekar, J., Ryba, N. J., and Zuker, C. S. 2000. A novel family of mammalian taste receptors. *Cell* 100:693–702.

Chandrashekar, J., Mueller, K. L., Hoon, M. A., Adler, E., Feng, L., Guo, W., Zuker, C. S., and Ryba, N. J. 2000. T2Rs function as bitter taste receptors. *Cell* 100:703–711.

Mano, I., and Driscoll, M. 1999. DEG/ENaC channels: A touchy superfamily that watches its salt. *BioEssays* 21:568–578.

Benos, D. J., and Stanton, B. A. 1999. Functional domains within the degenerin/epithelial sodium channel (Deg/ENaC) superfamily of ion channels. *J. Physiol. (Lond.)* 520(part 3):631–644.

McLaughlin, S. K., McKinnon, P. J., and Margolskee, R. F. 1992. Gustducin is a taste-cell-specific G protein closely related to the transducins. *Nature* 357:563–569.

Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J., and Zuker, C. S. 2001. Mammalian sweet taste receptors. *Cell* 106:381–390.

Vision

Stryer, L. 1988. Molecular basis of visual excitation. *Cold Spring Harbor Symp. Quant. Biol.* 53:283–294.

Jastrzebska, B., Tsybovsky, Y., and Palczewski, K. 2010. Complexes between photoactivated rhodopsin and transducin: Progress and questions. *Biochem. J.* 428:1–10.

Wald, G. 1968. The molecular basis of visual excitation. *Nature* 219:800–807.

Ames, J. B., Dizhoor, A. M., Ikura, M., Palczewski, K., and Stryer, L. 1999. Three-dimensional structure of guanylyl cyclase activating protein-2, a calcium-sensitive modulator of photoreceptor guanylyl cyclases. *J. Biol. Chem.* 274:19329–19337.

Nathans, J. 1994. In the eye of the beholder: Visual pigments and inherited variation in human vision. *Cell* 78:357–360.

Nathans, J. 1999. The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron* 24:299–312.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., LeTrong, I., Teller, D. C., Okada, T., Stenkamp, R. E., et al. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739–745.

Filipek, S, Teller, D. C., Palczewski, K., and Stemkamp, R. 2003. The crystallographic model of rhodopsin and its use in studies of other G protein-coupled receptors. *Annu. Rev. Biophys. Biomol. Struct.* 32:375–397.

Hearing

Furness, D. N., Hackney, C. M., and Evans, M. G. 2010. Localisation of the mechanotransducer channels in mammalian cochlear hair cells provides clues to their gating. *J. Physiol.* 588:765–772.

Lim, K., and Park, S. 2009. A mechanical model of the gating spring mechanism of stereocilia. *J. Biomech.* 42:2158–2164.

Siemens, J., Lillo, C., Dumont, R. A., Reynolds, A., Williams, D. S., Gillespie, P. G., and Muller, U. 2004. Cadherin 23 is a component of the tip link in hair-cell stereocilia. *Nature* 428:950–955.

Spinelli, K. J., and Gillespie, P. G. 2009. Bottoms up: Transduction channels at tip link bases. *Nat. Neurosci.* 12:529–530.

Hudspeth, A. J. 1997. How hearing happens. *Neuron* 19:947–950.

Pickles, J. O., and Corey, D. P. 1992. Mechanoelectrical transduction by hair cells. *Trends Neurosci.* 15:254–259.

Walker, R. G., Willingham, A. T., and Zuker, C. S. 2000. A *Drosophila* mechanosensory transduction channel. *Science* 287:2229–2234.

Hudspeth, A. J., Choe, Y., Mehta, A. D., and Martin, P. 2000. Putting ion channels to work: Mechanoelectrical transduction, adaptation, and amplification by hair cells. *Proc. Natl. Acad. Sci. U.S.A.* 97:11765–11772.

Touch and Pain Reception

Myers, B. R., Bohlen, C. J., and Julius, D. 2008. A yeast genetic screen reveals a critical role for the pore helix domain in TRP channel gating. *Neuron* 58:362–373.

Lishko, P. V., Procko, E., Jin, X., Phelps, C. B., and Gaudet, R. 2007. The ankyrin repeats of TRPV1 bind multiple ligands and modulate channel sensitivity. *Neuron* 54:905–918.

Franco-Obregon, A., and Clapham, D. E. 1998. Touch channels sense blood pressure. *Neuron* 21:1224–1226.

Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., and Julius, D. 1997. The capsaicin receptor: A heat-activated ion channel in the pain pathway. *Nature* 389:816–824.

Tominaga, M., Caterina, M. J., Malmberg, A. B., Rosen, T. A., Gilbert, H., Skinner, K., Raumann, B. E., Basbaum, A. I., and Julius, D. 1998. The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* 21:531–543.

Caterina, M. J., and Julius, D. 1999. Sense and specificity: A molecular identity for nociceptors. *Curr. Opin. Neurobiol.* 9:525–530.

Clapham, D. E. 2003. TRP channels as cellular sensors. *Nature* 426:517–524.

Chapter 34

Where to Start

Nossal, G. J. V. 1993. Life, death, and the immune system. *Sci. Am.* 269(3):53–62.

Tonegawa, S. 1985. The molecules of the immune system. *Sci. Am.* 253(4):122–131.

Leder, P. 1982. The genetics of antibody diversity. *Sci. Am.* 246(5):102–115.

Bromley, S. K., Burack, W. R., Johnson, K. G., Somersalo, K., Sims, T. N., Sumen, C., Davis, M. M., Shaw, A. S., Allen, P. M., and Dustin, M. L. 2001. The immunological synapse. *Annu. Rev. Immunol.* 19:375–396.

Books

Owen, J. A., Punt, J., and Stranford, S. A. 2013. *Kuby Immunology* (7th ed.). W. H. Freeman and Company.

Abbas, A. K., Lichtman, A. H. H., and Pillai, S. 2014. *Cellular and Molecular Immunology* (8th ed.). Saunders.

Cold Spring Harbor Symposia on Quantitative Biology. 1989. Volume 54. *Immunological Recognition.* Cold Spring Harbor Laboratory Press.

Weir, D. M. (Ed.). 1996. *Handbook of Experimental Immunology* (5th ed.). Oxford University Press.

Murphy, K. 2011. *Janeway's Immunobiology* (8th ed.). Garland Science.

Innate Immune System

Janeway, C. A., Jr., and Medzhitov, R. 2002. Innate immune recognition. *Annu. Rev. Immunol.* 20:197–216.

Khalturin, K., Panzer, Z., Cooper, M. D., and Bosch, T. C. 2004. Recognition strategies in the innate immune system of ancestral chordates. *Mol. Immunol.* 41:1077–1087.

Beutler, B., and Rietschel, E. T. 2003. Innate immune sensing and its roots: The story of endotoxin. *Nat. Rev. Immunol.* 3:169–176.

Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J. L., and Tong, L. 2000. Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 408:111–115.

Botos, I., Segal, D. M., and Davies, D. R. 2011. The structural biology of the Toll-like receptors. *Structure* 19:447–459.

Structure of Antibodies and Antibody-Antigen Complexes

Davies, D. R., Padlan, E. A., and Sheriff, S. 1990. Antibody-antigen complexes. *Annu. Rev. Biochem.* 59:439–473.

Poljak, R. J. 1991. Structure of antibodies and their complexes with antigens. *Mol. Immunol.* 28:1341–1345.

Davies, D. R., and Cohen, G. H. 1996. Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. U.S.A.* 93:7–12.

Marquart, M., Deisenhofer, J., Huber, R., and Palm, W. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.* 141:369–391.

Silverton, E. W., Navia, M. A., and Davies, D. R. 1977. Three-dimensional structure of an intact human immunoglobulin. *Proc. Natl. Acad. Sci. U.S.A.* 74:5140–5144.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith, G. S., and Davies, D. R. 1989. Structure of an antibody-antigen complex: Crystal structure of the HyHEL-10 Fab lysozyme complex. *Proc. Natl. Acad. Sci. U.S.A.* 86:5938–5942.

Rini, J., Schultze-Gahmen, U., and Wilson, I. A. 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* 255:959–965.

Fischmann, T. O., Bentley, G. A., Bhat, T. N., Boulot, G., Mariuzza, R. A., Phillips, S. E., Tello, D., and Poljak, R. J. 1991. Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution. *J. Biol. Chem.* 266:12915–12920.

Burton, D. R. 1990. Antibody: The flexible adaptor molecule. *Trends Biochem. Sci.* 15:64–69.

Saphire, E. O., Parren P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R., and Wilson, I. A. 2001. Crystal structure of a neutralizing human IgG against HIV-1: A template for vaccine design. *Science* 293:1155–1159.

Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., et al. 2003. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* 300:2065–2071.

Generation of Diversity

Tonegawa, S. 1988. Somatic generation of immune diversity. *Biosci. Rep.* 8:3–26.

Honjo, T., and Habu, S. 1985. Origin of immune diversity: Genetic variation and selection. *Annu. Rev. Biochem.* 54:803–830.

Gellert, M., and McBlane, J. F. 1995. Steps along the pathway of VDJ recombination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 347:43–47.

Harris, R. S., Kong, Q., and Maizels, N. 1999. Somatic hypermutation and the three R's: Repair, replication and recombination. *Mutat. Res.* 436:157–178.

Lewis, S. M., and Wu, G. E. 1997. The origins of V(D)J recombination. *Cell* 88:159–162.

Ramsden, D. A., van Gent, D. C., and Gellert, M. 1997. Specificity in V(D)J recombination: New lessons from biochemistry and genetics. *Curr. Opin. Immunol.* 9:114–120.

Roth, D. B., and Craig, N. L. 1998. VDJ recombination: A transposase goes to work. *Cell* 94:411–414.

Sadofsky, M. J. 2001. The RAG proteins in V(D)J recombination: More than just a nuclease. *Nucleic Acids Res.* 29:1399–1409.

MHC Proteins and Antigen Processing

Bjorkman, P. J., and Parham, P. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem.* 59:253–288.

Goldberg, A. L., and Rock, K. L. 1992. Proteolysis, proteasomes, and antigen presentation. *Nature* 357:375–379.

Madden, D. R., Gorga, J. C., Strominger, J. L., and Wiley, D. C. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight binding to MHC. *Cell* 70:1035–1048.

Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A., and Wilson, I. A. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* 257:880–881.

Matsumura, M., Fremont, D. H., Peterson, P. A., and Wilson, I. A. 1992. Emerging principles for the recognition of peptide antigens by MHC class I. *Science* 257:927–934.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L., and Wiley, D. C. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33–39.

Saper, M. A., Bjorkman, P. J., and Wiley, D. C. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* 219:277–319.

Madden, D. R., Gorga, J. C., Strominger, J. L., and Wiley, D. C. 1991. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* 353:321–325.

Cresswell, P., Bangia, N., Dick, T., and Diedrich, G. 1999. The nature of the MHC class I peptide loading complex. *Immunol. Rev.* 172:21–28.

Madden, D. R., Garboczi, D. N., and Wiley, D. C. 1993. The antigenic identity of peptide-MHC complexes: A comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* 75:693–708.

T-Cell Receptors and Signaling Complexes

Hennecke, J., and Wiley, D. C. 2001. T-cell receptor-MHC interactions up close. *Cell* 104:1–4.

Ding, Y. H., Smith, K. J., Garboczi, D. N., Utz, U., Biddison, W. E., and Wiley, D. C. 1998. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* 8:403–411.

Reinherz, E. L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y., Hussey, R. E., Smolyar, A., Hare, B., Zhang, R., et al. 1999. The crystal structure of a T-cell receptor in complex with peptide and MHC class II. *Science* 286:1913–1921.

Davis, M. M., and Bjorkman, P. J. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* 334:395–402.

Cochran, J. R., Cameron, T. O., and Stern, L. J. 2000. The relationship of MHC-peptide binding and T cell activation probed using chemically defined MHC class II oligomers. *Immunity* 12:241–250.

Garcia, K. C., Teyton, L., and Wilson, I. A. 1999. Structural basis of T cell recognition. *Annu. Rev. Immunol.* 17:369–397.

Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L. A., and Wilson, I. A. 1996. An αβ T-cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274:209–219.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E., Wiley, D. C. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384:134–141.

Gaul, B. S., Harrison, M. L., Geahlen, R. L., Burton, R. A., and Post, C. B. 2000. Substrate recognition by the Lyn protein-tyrosine kinase: NMR structure of the immunoreceptor tyrosine-based activation motif signaling region of the B cell antigen receptor. *J. Biol. Chem.* 275:16174–16182.

Kern, P. S., Teng, M. K., Smolyar, A., Liu, J. H., Liu, J., Hussey, R. E., Spoerl, R., Chang, H. C., Reinherz, E. L., and Wang, J. H. 1998. Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8 αβ ectodomain fragment in complex with H-2Kb. *Immunity* 9:519–530.

Konig, R., Fleury, S., and Germain, R. N. 1996. The structural basis of CD4-MHC class II interactions: Coreceptor contributions to T cell receptor antigen recognition and oligomerization-dependent signal transduction. *Curr. Top. Microbiol. Immunol.* 205:19–46.

Davis, M. M., Boniface, J. J., Reich, Z., Lyons, D., Hampl, J., Arden, B., and Chien, Y. 1998. Ligand recognition by αβ T-cell receptors. *Annu. Rev. Immunol.* 16:523–544.

Janeway, C. J. 1992. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu. Rev. Immunol.* 10:645–674.

Podack, E. R., and Kupfer, A. 1991. T-cell effector functions: Mechanisms for delivery of cytotoxicity and help. *Annu. Rev. Cell Biol.* 7:479–504.

Davis, M. M. 1990. T cell receptor gene diversity and selection. *Annu. Rev. Biochem.* 59:475–496.

Leahy, D. J., Axel, R., and Hendrickson, W. A. 1992. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. *Cell* 68:1145–1162.

Bots, M., and Medema, J. P. 2006. Granzymes at a glance. *J. Cell. Sci.* 119:5011–5014.

Lowin, B., Hahne, M., Mattmann, C., and Tschopp, J. 1994. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways. *Nature* 370:650–652.

Rudolph, M. G., and Wilson, I. A. 2002. The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* 14:52–65.

HIV and AIDS

Fauci, A. S. 1988. The human immunodeficiency virus: Infectivity and mechanisms of pathogenesis. *Science* 239:617–622.

Gallo, R. C., and Montagnier, L. 1988. AIDS in 1988. *Sci. Am.* 259(4): 41–48.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393:648–659.

Vaccines

Johnston, M. I. and Fauci, A. S. 2007. An HIV vaccine—evolving concepts. *New Engl. J. Med.* 356:2073–2081.

Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P., Nabel, G. J., Sodroski, J., Wilson, I. A., and Wyatt, R. T. 2004. HIV vaccine design and the neutralizing antibody problem. *Nature Immunol.* 5:233–236.

Ada, G. 2001. Vaccines and vaccination. *New Engl. J. Med.* 345: 1042–1053.

Behbehani, A. M. 1983. The smallpox story: Life and death of an old disease. *Microbiol. Rev.* 47:455–509.

Discovery of Major Concepts

Ada, G. L., and Nossal, G. 1987. The clonal selection theory. *Sci. Am.* 257(2):62–69.

Porter, R. R. 1973. Structural studies of immunoglobulins. *Science* 180:713–716.

Edelman, G. M. 1973. Antibody structure and molecular immunology. *Science* 180:830–840.

Kohler, G. 1986. Derivation and diversification of monoclonal antibodies. *Science* 233:1281–1286.

Milstein, C. 1986. From antibody structure to immunological diversification of immune response. *Science* 231:1261–1268.

Janeway, C. A., Jr. 1989. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symp. Quant. Biol.* 54:1–13.

Jerne, N. K. 1971. Somatic generation of immune recognition. *Eur. J. Immunol.* 1:1–9.

Chapter 35

Where to Start

Gennerich, A., and Vale, R. D. 2009. Walking the walk: How kinesin and dynein coordinate their steps. *Curr. Opin. Cell Biol.* 21:59–67.

Vale, R. D. 2003. The molecular motor toolbox for intracellular transport. *Cell* 112:467–480.

Vale, R. D., and Milligan, R. A. 2000. The way things move: Looking under the hood of molecular motor proteins. *Science* 288:88–95.

Vale, R. D. 1996. Switches, latches, and amplifiers: Common themes of G proteins and molecular motors. *J. Cell Biol.* 135:291–302.

Mehta, A. D., Rief, M., Spudich, J. A., Smith, D. A., and Simmons, R. M. 1999. Single-molecule biomechanics with optical methods. *Science* 283:1689–1695.

Schuster, S. C., and Khan, S. 1994. The bacterial flagellar motor. *Annu. Rev. Biophys. Biomol. Struct.* 23:509–539.

Books

Howard, J. 2001. *Mechanics of Motor Proteins and the Cytoskeleton.* Sinauer.

Squire, J. M. 1986. *Muscle Design, Diversity, and Disease.* Benjamin Cummings.

Pollack, G. H., and Sugi, H. (Eds.). 1984. *Contractile Mechanisms in Muscle.* Plenum.

Myosin and Actin

Lorenz, M., and Holmes, K. C. 2010. The actin-myosin interface. *Proc. Natl. Acad. Sci. U.S.A.* 107:12529–12534.

Yang, Y., Gourinath, S., Kovacs, M., Nyitray, L., Reutzel, R., Himmel, D. M., O'Neall-Hennessey, E., Reshetnikova, L., Szent-Györgyi, A. G., Brown, J. H., et al. 2007. Rigor-like structures from muscle myosins reveal key mechanical elements in the transduction pathways of this allosteric motor. *Structure* 15:553–564.

Himmel, D. M., Mui, S., O'Neall-Hennessey, E., Szent-Györgyi, A. G., and Cohen, C. 2009. The on-off switch in regulated myosins: Different triggers but related mechanisms. *J. Mol. Biol.* 394: 496–505.

Houdusse, A., Gaucher, J. F., Krementsova, E., Mui, S., Trybus, K. M., and Cohen, C. 2006. Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. *Proc. Natl. Acad. Sci. U.S.A.* 103:19326–19331.

Li, X. E., Holmes, K. C., Lehman, W., Jung, H., and Fischer, S. 2010. The shape and flexibility of tropomyosin coiled coils: Implications for actin filament assembly and regulation. *J. Mol. Biol.* 395: 327–339.

Fischer, S., Windshugel, B., Horak, D., Holmes, K. C., and Smith, J. C. 2005. Structural mechanism of the recovery stroke in the myosin molecular motor. *Proc. Natl. Acad. Sci. U.S.A.* 102:6873–6878.

Holmes, K. C., Angert, I., Kull, F. J., Jahn, W., and Schroder, R. R. 2003. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* 425:423–427.

Holmes, K. C., Schroder, R. R., Sweeney, H. L., and Houdusse, A. 2004. The structure of the rigor complex and its implications for the power stroke. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359: 1819–1828.

Purcell, T. J., Morris, C., Spudich, J. A., and Sweeney, H. L. 2002. Role of the lever arm in the processive stepping of myosin V. *Proc. Natl. Acad. Sci. U.S.A.* 99:14159–14164.

Purcell, T. J., Sweeney, H. L., and Spudich, J. A. 2005. A force-dependent state controls the coordination of processive myosin V. *Proc. Natl. Acad. Sci. U.S.A.* 102:13873–13878.

Holmes, K. C. 1997. The swinging lever-arm hypothesis of muscle contraction. *Curr. Biol.* 7:R112–R118.

Berg, J. S., Powell, B. C., and Cheney, R. E. 2001. A millennial myosin census. *Mol. Biol. Cell* 12:780–794.

Houdusse, A., Kalabokis, V. N., Himmel, D., Szent-Györgyi, A. G., and Cohen, C. 1999. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: A novel conformation of the myosin head. *Cell* 97:459–470.

Houdusse, A., Szent-Györgyi, A. G., and Cohen, C. 2000. Three conformational states of scallop myosin S1. *Proc. Natl. Acad. Sci. U.S.A.* 97:11238–11243.

Uyeda, T. Q., Abramson, P. D., and Spudich, J. A. 1996. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl. Acad. Sci. U.S.A.* 93:4459–4464.

Mehta, A. D., Rock, R. S., Rief, M., Spudich, J. A., Mooseker, M. S., and Cheney, R. E. 1999. Myosin-V is a processive actin-based motor. *Nature* 400:590–593.

Otterbein, L. R., Graceffa, P., and Dominguez, R. 2001. The crystal structure of uncomplexed actin in the ADP state. *Science* 293:708–711.

Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. 1990. Atomic model of the actin filament. *Nature* 347:44–49.

Schutt, C. E., Myslik, J. C., Rozycki, M. D., Goonesekere, N. C., and Lindberg, U. 1993. The structure of crystalline profilin-β-actin. *Nature* 365:810–816.

van den Ent, F., Amos, L. A., and Lowe, J. 2001. Prokaryotic origin of the actin cytoskeleton. *Nature* 413:39–44.

Schutt, C. E., and Lindberg, U. 1998. Muscle contraction as a Markov process I: Energetics of the process. *Acta Physiol. Scand.* 163:307–323.

Rief, M., Rock, R. S., Mehta, A. D., Mooseker, M. S., Cheney, R. E., and Spudich, J. A. 2000. Myosin-V stepping kinetics: A molecular model for processivity. *Proc. Natl. Acad. Sci. U.S.A.* 97:9482–9486.

Friedman, T. B., Sellers, J. R., and Avraham, K. B. 1999. Unconventional myosins and the genetics of hearing loss. *Am. J. Med. Genet.* 89:147–157.

Kinesin, Dynein, and Microtubules

Cho, C. and Vale, R. D. 2012. The mechanism of dynein motility: Insight from crystal structures of the motor domain. *Biochim. Biophys. Acta* 1823:182–191.

Yildiz, A., Tomishige, M., Gennerich, A., and Vale, R. D. 2008. Intramolecular strain coordinates kinesin stepping behavior along microtubules. *Cell* 134:1030–1041.

Yildiz, A., Tomishige, M., Vale, R. D., and Selvin, P. R. 2004. Kinesin walks hand-over-hand. *Science* 303:676–678.

Rogers, G. C., Rogers, S. L., Schwimmer, T. A., Ems-McClung, S. C., Walczak, C. E., Vale, R. D., Scholey, J. M., and Sharp, D. J. 2004. Two mitotic kinesins cooperate to drive sister chromatid separation during anaphase. *Nature* 427:364–370.

Vale, R. D., and Fletterick, R. J. 1997. The design plan of kinesin motors. *Annu. Rev. Cell. Dev. Biol.* 13:745–777.

Kull, F. J., Sablin, E. P., Lau, R., Fletterick, R. J., and Vale, R. D. 1996. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* 380:550–555.

Kikkawa, M., Sablin, E. P., Okada, Y., Yajima, H., Fletterick, R. J., and Hirokawa, N. 2001. Switch-based mechanism of kinesin motors. *Nature* 411:439–445.

Wade, R. H., and Kozielski, F. 2000. Structural links to kinesin directionality and movement. *Nat. Struct. Biol.* 7:456–460.

Yun, M., Zhang, X., Park, C. G., Park, H. W., and Endow, S. A. 2001. A structural pathway for activation of the kinesin motor ATPase. *EMBO J.* 20:2611–2618.

Kozielski, F., De Bonis, S., Burmeister, W. P., Cohen-Addad, C., and Wade, R. H. 1999. The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations. *Struct. Fold. Des.* 7:1407–1416.

Lowe, J., Li, H., Downing, K. H., and Nogales, E. 2001. Refined structure of αβ-tubulin at 3.5 Å resolution. *J. Mol. Biol.* 313:1045–1057.

Nogales, E., Downing, K. H., Amos, L. A., and Lowe, J. 1998. Tubulin and FtsZ form a distinct family of GTPases. *Nat. Struct. Biol.* 5:451–458.

Zhao, C., Takita, J., Tanaka, Y., Setou, M., Nakagawa, T., Takeda, S., Yang, H. W., Terada, S., Nakata, T., Takei, Y., et al. 2001. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bb. *Cell* 105:587–597.

Asai, D. J., and Koonce, M. P. 2001. The dynein heavy chain: Structure, mechanics and evolution. *Trends Cell Biol.* 11:196–202.

Mocz, G., and Gibbons, I. R. 2001. Model for the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases. *Structure* 9:93–103.

Bacterial Motion and Chemotaxis

Baker, M. D., Wolanin, P. M., and Stock, J. B. 2006. Systems biology of bacterial chemotaxis. *Curr. Opin. Microbiol.* 9:187–192.

Wolanin, P. M., Baker, M. D., Francis, N. R., Thomas, D. R., DeRosier, D. J., and Stock, J. B. 2006. Self-assembly of receptor/signaling complexes in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U.S.A.* 103:14313–14318.

Sowa, Y., Rowe, A. D., Leake, M. C., Yakushi, T., Homma, M., Ishijima, A., and Berry, R. M. 2005. Direct observation of steps in rotation of the bacterial flagellar motor. *Nature* 437:916–919.

Berg, H. C. 2000. Constraints on models for the flagellar rotary motor. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355:491–501.

DeRosier, D. J. 1998. The turn of the screw: The bacterial flagellar motor. *Cell* 93:17–20.

Ryu, W. S., Berry, R. M., and Berg, H. C. 2000. Torque-generating units of the flagellar motor of *Escherichia coli* have a high duty ratio. *Nature* 403:444–447.

Lloyd, S. A., Whitby, F. G., Blair, D. F., and Hill, C. P. 1999. Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. *Nature* 400:472–475.

Purcell, E. M. 1977. Life at low Reynolds number. *Am. J. Physiol.* 45:3–11.

Macnab, R. M., and Parkinson, J. S. 1991. Genetic analysis of the bacterial flagellum. *Trends Genet.* 7:196–200.

Historical Aspects

Huxley, H. E. 1965. The mechanism of muscular contraction. *Sci. Am.* 213(6):18–27.

Summers, K. E., and Gibbons, I. R. 1971. ATP-induced sliding of tubules in trypsin-treated flagella of sea-urchin sperm. *Proc. Natl. Acad. Sci. U.S.A.* 68:3092–3096.

Macnab, R. M., and Koshland, D. E., Jr. 1972. The gradient-sensing mechanism in bacterial chemotaxis. *Proc. Natl. Acad. Sci. U.S.A.* 69:2509–2512.

Taylor, E. W. 2001. 1999 E. B. Wilson lecture: The cell as molecular machine. *Mol. Biol. Cell* 12:251–254.

Chapter 36

Where to Start

Gilman, A. G. 2012. Silver spoons and other personal reflections. *Annu. Rev. Pharmacol. Toxicol.* 52:1–19.

Zhang, H.-Y., Chen, L.-L., Xue-Juan Li, X.-J. and Zhang, J. 2010. Evolutionary inspirations for drug discovery. *Trends Pharmacol. Sci.* 31:443–448.

Books

Kenakin, T. P. 2014. *A Pharmacology Primer: Techniques for More Effective and Strategic Drug Discovery* (4th ed.). Academic Press.

Brunton, L., Chabner, B., and Knollman, B. 2011. *Goodman and Gilman's The Pharmacological Basis of Therapeutics* (12th ed.). McGraw-Hill Professional.

Walsh, C. T., and Schwartz-Bloom, R. D. 2004. *Levine's Pharmacology: Drug Actions and Reactions* (7th ed.). Taylor and Francis Group.

Silverman, R. B., and Holladay, M. W. 2014. *Organic Chemistry of Drug Design and Drug Action* (3d ed.). Academic Press.

Walsh, C. 2003. *Antibiotics: Actions, Origins, Resistance.* ASM Press.

ADME and Toxicity

Caldwell, J., Gardner, I., and Swales, N. 1995. An introduction to drug disposition: The basic principles of absorption, distribution, metabolism, and excretion. *Toxicol. Pathol.* 23:102–114.

Lee, W., and Kim, R. B. 2004. Transporters and renal drug elimination. *Annu. Rev. Pharmacol. Toxicol.* 44:137–166.

Lin, J., Sahakian, D. C., de Morais, S. M., Xu, J. J., Polzer, R. J., and Winter, S. M. 2003. The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr. Top. Med. Chem.* 3:1125–1154.

Poggesi, I. 2004. Predicting human pharmacokinetics from preclinical data. *Curr. Opin. Drug Discov. Devel.* 7:100–111.

Case Histories

Flower, R. J. 2003. The development of COX2 inhibitors. *Nat. Rev. Drug Discov.* 2:179–191.

Tobert, J. A. 2003. Lovastatin and beyond: The history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* 2:517–526.

Vacca, J. P., Dorsey, B. D., Schleif, W. A., Levin, R. B., McDaniel, S. L., Darke, P. L., Zugay, J., Quintero, J. C., Blahy, O. M., Roth, E., et al. 1994. L-735,524: An orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 91:4096–4100.

Wong, S., and Witte, O. N. 2004. The BCR-ABL story: Bench to bedside and back. *Annu. Rev. Immunol.* 22:247–306.

Structure-Based Drug Design

Kuntz, I. D. 1992. Structure-based strategies for drug design and discovery. *Science* 257:1078–1082.

Dorsey, B. D., Levin, R. B., McDaniel, S. L., Vacca, J. P., Guare, J. P., Darke, P. L., Zugay, J. A., Emini, E. A., Schleif, W. A., Quintero, J. C., et al. 1994. L-735,524: The design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.* 37:3443–3451.

Chen, Z., Li, Y., Chen, E., Hall, D. L., Darke, P. L., Culberson, C., Shafer, J. A., and Kuo, L. C. 1994. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* 269:26344–26348.

Combinatorial Chemistry

Baldwin, J. J. 1996. Design, synthesis and use of binary encoded synthetic chemical libraries. *Mol. Divers.* 2:81–88.

Burke, M. D., Berger, E. M., and Schreiber, S. L. 2003. Generating diverse skeletons of small molecules combinatorially. *Science* 302:613–618.

Edwards, P. J., and Morrell, A. I. 2002. Solid-phase compound library synthesis in drug design and development. *Curr. Opin. Drug Discov. Devel.* 5:594–605.

Genomics

Zambrowicz, B. P., and Sands, A. T. 2003. Knockouts model the 100 best-selling drugs: Will they model the next 100? *Nat. Rev. Drug Discov.* 2:38–51.

Salemme, F. R. 2003. Chemical genomics as an emerging paradigm for postgenomic drug discovery. *Pharmacogenomics* 4:257–267.

Michelson, S., and Joho, K. 2000. Drug discovery, drug development and the emerging world of pharmacogenomics: Prospecting for information in a data-rich landscape. *Curr. Opin. Mol. Ther.* 2:651–654.

Weinshilboum, R., and Wang, L. 2004. Pharmacogenomics: Bench to bedside. *Nat. Rev. Drug Discov.* 3:739–748.

INDEX

Note: Page numbers followed by f, t, and b refer to figures, tables, and boxed material.
Boldface page numbers indicate structural formulas and ribbon diagrams.

A band, 1020f
A site, ribosomal, 905, 905f
AAA ATPases
  in amino acid degradation, 685
  in dynein, 1012, 1014
ABC transporters
  ATP hydrolysis reaction and, 376
  defined, 367, 374
  domain arrangement of, 374, 374f
  in drug resistance, 1054
  eukaryotic, 376
  mechanism, 375, 375f
  structure of, **375**
ABCA1 (ATP-binding cassette transporter, subfamily A1), 787
ABO blood groups, 331–332, 331f
Abortive complex, 688
Absorption
  of dietary proteins, 682, 683f
  drug, 1036–1037, 1037f
  light, 569, 569f, 971–972
Abzymes, production of, 241
Acceptor control. *See* Respiratory control
Accessory pigments, in photosynthesis, 581–584, 582f, 583f
Acetaminophen, 1040–1041, 1040f
Acetoacetate. *See also* Ketone bodies
  conversion into acetyl CoA, 658–659
  from phenylalanine, 704
  in respiration, 658
  from tryptophan, 704
  utilization as fuel, 659, 659f
Acetoacetyl CoA, 698–699
Acetyl ACP, 663
Acetyl carnitine, 649–650
Acetyl CoA (acetyl coenzyme A)
  acetoacetate conversion into, 658–659
  acetyl-group-transfer potential, **437**
  in amino acid degradation, 698–699, 701–703
  biotin and, 480
  in citric acid cycle, 495, 498–500, 501–503, 503f, 515
  defined, 437
  in fatty acid metabolism, 648–649
  in fatty acid oxidation, 650–652
  in fatty acid synthesis, 512
  formation from ketone bodies, 657–658, 658f
  formation of, 464, 499
  in glyoxylate cycle, 516–517, 517f
  pyruvate dehydrogenase complex and, 499
  structure of, 437, **437**
  synthesis of, 498–500
  transfer to cytoplasm, 666–667, 666f
Acetyl CoA carboxylase
  in citrate concentration, 671, 672f
  in fatty acid metabolism, 662, 662t, 670–671
  filaments of, 671f
  regulation by hormones, 671–672
  regulation of, 671, 671f
Acetyl CoA synthetase, 649
Acetyl groups, 57
Acetyl transacylase, 663, 663t
Acetylation
  defined, 294t
  in gene regulation, 952–953, 954
  in metabolism regulation, 294
  protein, 293–294
Acetylcholine, 385, 386
Acetylcholine receptor
  defined, 385
  as ligand-gated channel, 386
  as nonspecific cation channel, 388
  opening, 387, 387f
  structure of, 386, 386f
N-acetylglucosamine phosphotransferase, 332
N-acetylglutamate, 694
N-acetylglutamate synthase, 694
Acetyllysine-binding domain, 953, 953f
N-acetylneuraminate, 772
N-acetyl-L-phenylalanine *p*-nitrophenyl ester, 254, 255f
Acid-base reactions
  in biological processes, 13–14
  defined, 13
  double helix and, 14–15
  hydrogen ions in, 13
Acids. *See also* Amino acids; Fatty acids; Nucleic acids
  bile, 646, 646f
  protonation/deprotonation of, 15–16
  sialic, 772
Aconitase, 504, 504f, 955, 956f
Acquired immune deficiency syndrome (AIDS). *See* Human immunodeficiency virus (HIV) infection
Actin/actin filaments
  barbed, 1017
  decorated, 1018
  defined, 1016
  F, 1017
  formation of, 1017–1018
  G, 1017
  in muscle contraction, 1019–1021
  myosin and, 1018–1022, 1020f
  pointed, 1017
  polymerization of, 1018
  structure of, 179, **179,** 1016–1017, 1017f
  in thin filaments, 1020
Actinomycin, 869, 870f
Action potentials
  defined, 378
  integration, 387–388
  ion channels and, 378, 378f
  mechanism, 388, 388f
Activated carriers
  as derived from vitamins, 438–440
  of electrons, for fuel oxidation, 435–436
  of electrons, for reductive biosynthesis, 436
  in metabolism, 435–440, 437t
  recurring set of, 437–438
  of two-carbon fragments, 437–438
Activated glycoaldehyde unit, 606
Activated methyl cycle, 725, 725f
Activation domains, 945, 946
Activation energy
  defined, 221
  enzyme decrease of, 221, 221f, 222
  symbol, 221
Active membrane transport, 367, 369
Active sites
  ATCase, 289, 289f, **289**
  chymotrypsin, 255–256, 256f
  concentration of, 230
  as crevice, 223
  defined, 216
  enzyme and substrate interaction, 223
  evolution of, 252
  as microenvironments, 224
  myosin, 279, 279f
  proteases, 261f
  residues, 223, 223f
  RNA polymerase, 861
  volume of enzyme, 223–224
Acute intermittent porphyria, 738
Acyclovir monophosphate, 748
Acyl carrier protein (ACP), 661–662, 662f, **662,** 664–665, 665f
Acyl CoA:cholesterol acyltransferase (ACAT), 784
Acyl intermediates, 239, 240f
Acyl-enzyme intermediates, 255, 255f, 686
Adaptive immune system, 984–985
Adaptor proteins, 410
Adenine, **4,** 5f, 110, **110**
  deamination of, 846, 846f
  methylation of, 274–275, 275f
  resonance structures, 7
Adenine phosphoribosyltransferase, 752
Adenosine, 107
Adenosine 5′-triphosphate. *See* ATP
Adenosine deaminase, 164, 760–761
Adenosine diphosphate. *See* ADP
Adenosine triphosphate. *See* ATP
S-adenosylhomocysteine, 725, **725**
S-adenosylmethionine (SAM)
  in activated methyl cycle, 725, 725f
  in adenylation of methionine, 701

S-adenosylmethionine (SAM)—(*continued*)
  in amino acid degradation, 701, **702**
  in amino acid synthesis, 724–726
  as ethylene precursor, 726
  in phospholipid synthesis, 771
Adenylate, 108, 751
Adenylate cyclase, 402–403, 402f
Adenylate kinase, **280**
Adenylated intermediates, 721
Adenylation
  in amino acid activation, 898–899
  in amino acid synthesis, 732–733, 733f
  in translation, 899–902
Adenylosuccinate synthetase, 751
Adenylyl transferase, 733, 733f
Adipokines, 805
Adiponectin, 806
Adipose cells (fat cells), 645
Adipose tissue
  brown, 555, 555f
  fat in, 815
  glucose entry into, 816
  glycerol released from, 817
  triacylglycerols in, 818
  white, 555
Adipose triglyceride lipase (ATGL), 647–648, 647f
ADME properties, of drugs, 1036–1040, 1037f
ADP (adenosine diphosphate), **275**
  ATP hydrolyzed to, 426
  in citric acid cycle, 406f, 506
  formation of, 275
  in kinesin movement, 1025, 1025f
  in oxidative phosphorylation, 550–551, 551f, 553–554, 554f
  in random sequential reaction, 232
  structure of, 426, **427**
Adrenocorticotropic hormone, 59
Aerobic glycolysis, 474–476, 510
Affinity chromatography, 70–71, 70f
Affinity maturation, 993
Affinity tags, 78, 237–238, 237f, 238f
Aflatoxin activation, 846f
A-form DNA, 111–112, 111f, 112f, 112t
Aggrecan, 328
α-glucosidase, 450
Agonists, 951
Agouti-related peptide (AgRP), 805–806, 806f
Agre, Peter, 1
Agrobacterium tumefaciens, 163, 163f
Alanine
  defined, 30
  formation of, 720–721
  in gluconeogenesis, 486
  in muscle contraction, 486, 486f
  structure of, 31f
  tryptophan in, 700
Alanine aminotransferase, 688
Alanyl-tRNA, 895, 895f
Albinism, 705t
Albumin, 137f, 1037
Alcaptonuria, 705
Alcohol dehydrogenase, 820
Alcoholic fermentation, 449, 463, 463f
Alcoholic scurvy, 821
Alcohols
  fermentation of, 462–464, 463f
  formation from pyruvate, 462–464, 463f
  metabolism of, 819–822
  monosaccharides and, 322
  toxicity of, 821–822
Aldehyde dehydrogenase, 228
Aldehydes, oxidation of energy, 457–458
Alder, Julius, 1026
Aldimine, 690
Aldolase, 455
  in Calvin cycle, 595
  in transaldolase reaction, 606–607
Aldoses
  conversion to ketoses, 454
  defined, 316
Aldosterone, 789, 792, **793**
Allolactose, 930
Allopurinol, 761
Allosteric constant, 290
Allosteric control, 285
Allosteric effector, 200–201
Allosteric enzymes, 233, 233f
  allosteric constant, 290
  concerted model, 290
  control of, 443
  homotropic effects on, 290
  sequential model, 290–291
  sigmoidal kinetics, 287, 287f
  substrate concentration and, 291
  threshold effects, 291, 291f
  T-to-R equilibrium in, 291–292, 291f, 292f
  T-to-R state transition, 290–291, 290f
Allosteric sites, 287
Alloxanthine, 761
α helix
  of membrane proteins, 355–356, 355f, 355t
  pitch of, 40
  as right-handed, 41
  schematic view of, 41f
  screw sense, 40–41, 40f
  structure of, 41f, **41**
α-stabilizing protein (AHSP), 207–208, 207f
α_1-antiproteinase, 303
α-amino acids, 29, 35
α-helical coiled coil, 44
α-hemoglobin, 207, 207f
  alignment with gap insertion, 173f
  amino acid sequences, 171–172, 171f, 172f
α-thalassemia, 207
Alternative splicing, 128, 129f, 877
  defects in, 885t
  defined, 885
  example of, 885f
  pre-mRNA, 885–886
Altman, Sidney, 859, 870
Alu sequences, 154
Alzheimer's disease, 56–57
Amanita phalloides, 873, 873f
α-amanitin, 872–873
Ames test, 852, 852f
Amide bonds. *See* Peptide bonds
Amiloride-sensitive sodium channel, 969, 969f, 977
Amines, monosaccharides and, 322
Amino acid degradation, 681–708
  acetoacetyl CoA in, 698–699
  acetyl CoA in, 698–699, 701–703
  amino acid use through, 681
  aminotransferases in, 688–691
  of aromatic amino acids, 703–705
  of branched-chain amino acids, 687–689, 701–703
  coenzymes in, 682
  deamination in, 692
  digestive enzymes in, 682
  fumarate in, 698
  glucose-alanine cycle in, 692–693, 693f
  glutamate dehydrogenase in, 688–689
  α-ketobutyrate in, 701
  α-ketoglutarate in, 698, 700–701, 700f
  liver in, 687, 692–693
  metabolism errors and, 705–706, 705t
  nitrogen removal in, 687–693
  overview of, 681–682
  oxaloacetate in, 698, 700
  oxygenases in, 703–705
  propionyl CoA in, 701–702
  proteasomes in, 681, 685–686, 686f
  in protein turnover, 683–687
  pyridoxal phosphate (PLP) in, 682, 691–692
  pyruvate in, 692–693, 698, 699–700, 699f
  regulation function of, 683–687, 687t
  Schiff-base intermediates in, 689–690
  serine dehydratase in, 692
  succinyl CoA in, 698, 701, 701f
  threonine dehydratase in, 692
  transaminases in, 688–691
  ubiquitin in, 683–685, 684f
  ubiquitination in, 684–686
  urea cycle in, 693–698
Amino acid sequences
  alignment of, 171–177, 172f, 173f
  collagen, 45, 45f
  comparison of, 90
  determination of, 65, 66
  directionality of, 35–36, 36f
  in DNA probe creation, 91
  DNA sequences and, 90f
  evolution of, 37
  gap insertion, 172–173, 173f
  of hemoglobin, 171–172, 171f, 172f
  homologous, 177
  identities, 172
  importance of, 37
  information provided by, 90–91
  of insulin, 37f
  of myoglobin, 171–172, 171f, 172f
  porin, 352, 352f

  probes generated from, 147, 147f
  in protein identification, 90–91
  in protein structure determination, 49–59
  protein structures and, 37, 37f
  recombinant DNA technology and, 89–90
  ribonuclease, 49f
  searching for internal repeats, 91, 91f
  shuffled, 173–174, 173f
  as signals, 91
  statistical comparison of alignment, 173, 173f
  substitution in, 174–177, 175f, 176f
  transmembrane helices from, 354–356, 355f
Amino acid side chains
  aliphatic, 32
  aromatic, 32
  charge of, 33–35, 33f, 34f
  covalent modifications, 57f
  defined, 29
  distribution of, 46
  hydrocarbon, 32
  hydrophilic, 33
  hydrophobic, 32
  hydroxyl-containing, 32, 32f
  ionizable, 33–34, 34t
  sulfydryl-containing, 33, 33f
  thiol-containing, 33, 33f
  types of, 30
Amino acid synthesis, 713–739
  S-adenosylmethionine in, 724–726
  adenylation in, 732–733, 733f
  ammonia in, 714–718
  aromatic amino acids, 727–729
  branched pathways in, 731–732
  chorismate in, 727–729, 727f
  enzyme multiplicity in, 732, 732f
  essential amino acids, 719–720, 720f
  feedback inhibition/activation in, 731–732
  glutamine synthetase in, 732–733, 733f
  human, 719–720
  α-ketoglutarate in, 717–718, 720
  nitrogen fixation in, 714–718, 715f
  nonessential amino acids, 719, 720f
  overview of, 713–714
  regulation of, 730–733
  regulatory domains in, 731–732, 732f
  shikimate in, 727–729
  tetrahydrofolate in, 723–724, 723f, 723t, **724**
  threonine deaminase in, 731–732, 731f
  by transamination, 720–721, 720f
Amino acids. *See also* Proteins
  abbreviations for, 34–35, 34t
  acetyl groups, 57
  activation by adenylation, 898–899
  activation sites, 899–900, 900f
  aromatic, 32, 703–705, 727–729
  basic set of, 719t
  branched-chain, 687–689, 701–703
  carbon skeletons of, fates of, 699f
  chirality of, 29, 721, 721f
  code words, 125
  D isomer, 29, 29f
  defined, 21
  dipolar form of, 29–30, 30f
  encoding of, 124–127
  essential, 682, 682t, 719, 720f
  excess, 681
  genetic code, 125, 125t
  glucogenic, 698–703
  in gluconeogenesis, 476
  hydrophilic, 33, 47
  hydrophobic, 30–32, 31f
  hydroxyl groups of, 32–33, 57
  ionization state of, 29–30, 30f
  L isomer, 29, 29f
  metabolism, 705–706, 705t
  negatively charged, 33–34, 34f
  nitrogen in, 714
  nonessential, 719, 720f
  peptide bonds, 35–36, 35f
  polar, 32–33, 32f
  positively charged, 33, 33f
  as precursors, 734–738
  pyruvate formation from, 699, 699f
  reactive, 33
  residue, 35, 47
  in secondary structures, 51t
  tRNA attachment, 123, 123f
  undesirable reactivity in, 35f
Amino-acid signal sequences, 916–918, 916f
Aminoacrylate, Schiff bases of, 729, **729**
Aminoacyl adenylate, 898
Aminoacyl-AMP, 899
Aminoacyl-tRNA, 123, 124f, 896–897, 897f
  defined, 898
  editing of, 900–901, 900f
  elongation factors and, 907–908, 907f
Aminoacyl-tRNA synthetases
  accuracy of, 901
  activation sites of, 899–900, 900f
  amino acid activation sites, 899–900
  classes of, 901–902, 902f
  defined, 894, 898
  double sieve in, 901
  editing sites, 900–901, 900f
  evolution of, 902
  genetic code and, 898–902
  proofreading by, 900–901
  subunit structure, 902, 902t
Aminoglycoside antibiotics, 914
δ-aminolevulinate, 736
δ-aminolevulinate synthase, 736
Aminopeptidase N, 682
Aminopterin, 757
Aminotransferases
  in amino acid degradation, 688–691
  blood levels of, 691
  defined, 688
  Schiff-base intermediates formation in, 689–690
Ammonia
  in amino acid degradation, 687–689
  in amino acid synthesis, 714–718
  conversion to urea, 693–698
  formation of, 687–689
  in glutamate synthesis, 717–718
  in glutamine synthesis, 717–718
  in pyrimidine synthesis, 745
  transport of, 692–693
Ammoniotelic organisms, 698
AMP (adenosine monophosphate)
  ATP hydrolyzed to, 426
  energy charge and, 482
  in purine synthesis, 751–752, 751f
  structure of, 426, **427**
  synthesis of, 758–759, 758f
AMP-activated protein kinase (AMPK), 671, 806, 812
Amphibolic pathways, 425
Amycolatopsis, 869
α-amylase, 324, 450
Amyloid fibers, 56–57, 56f
Amyloid plaques, 56f
Amyloid precursor protein (APP), 57
Amyloidoses, 56–57
Amylopectin, 324
Amyotrophic lateral sclerosis (ALS), 136
Amytal, 556
Anabolic steroids, 951
Anabolism, 424, 714b
Analytes, 85
Anastrozole, 793
Anchor residues, 997–998, 997f
Anderson disease, 637t
Androgens
  functions of, 789
  pathways for formation of, 794f
  synthesis of, 792–793, **794**
Androstenedione, 789, 793, **794**
Anemia, 17–18, 205–206, 207
Anfinsen, Christian, 49
Angelman syndrome, 685
Angiogenin, sequence comparison of, **170**
Animal models, in drug target testing, 1049
Animal testing, in drug development, 1044
Ankyrin repeats, 976–977, 976f
Anomers, 317f, 319
Anosmias, 962
Antagonists, 952
Anthranilate, 728
Anthrax, 839
Antibiotics
  aminoglycoside, 914
  protein synthesis inhibition by, 913–915, 914t
  transcription inhibition by, 869–870, 869f
  translocation inhibition by, 914–915, 915f
Antibodies. *See also* Immunoglobulins
  amino acid sequence data in making, 91
  antibody binding of, 989–990, 989f
  antigen interaction, 80f
  antigen-binding specificity, 995
  class switching in, 994–995, 995f
  constant region of, 988, 988f
  defined, 79, 984
  diversity of, 992–993

- Antibodies. *See also* Immunoglobulins—(*continued*)
  - formation of, 994–995
  - hypervariable loops of, 988–990, 988f
  - monoclonal, 80–82, 80f, 81f
  - number of, 992
  - oligomerization of, 993–994
  - polyclonal, 80, 80f
  - primary, 83–84
  - recombination in, 991–995, 991f, 992f
  - secondary, 84
  - structures of, 79f, 985–986
  - variable region of, 988, 988f, 989f
- Antibody-protein interactions, 990, 990f
- Anticancer drugs
  - in blocking thymidylate synthesis, 757–758
  - targets, 757f
- Anticoagulants, 307
- Anticodon loop, 896
- Anticodons
  - base/base pair, 898, 898f
  - defined, 123
  - in translation, 896, 898, 898f, 901
- Antidiuretic hormone, 92–93, 93f
- Antigenic determinants, 79
- Antigens
  - ABO blood group, 331–332, 331f
  - antibody interaction, 80f
  - cross-linking of, 986, 986f
  - defined, 79
  - synthetic peptides as, 92
- Antihemophilic factor, 308
- Antiporters, 376, 376f
- Antiserum, 80
- Antithrombin III, 307–308
- AP endonuclease, 849
- Apolipoprotein B (apo B), 879, 879f
- Apolipoproteins, 646
- Apoptosis
  - defined, 299, 557
  - mitochondria in, 557–558
- Apoptosome, 558
- Apoptotic peptidase-activating factor 1 (APAF-1), 558
- Approximation, catalysis by, 252
- Aptamers, 187
- Aquaporins, 391, 391f
- *Arabidopsis thaliana*, 154
- Arachidonate, 669, 669f
- Archaea, 3, 3f
  - membranes of, 346–347, 346f
  - proteasomes of, 687, 687f
- Arginase, 695
- Arginine
  - argininosuccinase and, 695
  - degradation, 701, 701f
  - structure of, 33, 33f
  - synthesis of, 722, 722f
- Argininosuccinase, 695
- Argininosuccinase deficiency, 697, 697f
- Argininosuccinate, 695, 696
- Argininosuccinate synthetase, 695
- Argonaute complex, 956, 956f
- Aromatase, 793
- Aromatase inhibitors, 793
- Aromatic amino acids
  - degradation of, 703–705
  - side chains, 32
  - synthesis of, 727–729
- Arrestin, 972
- Arsenite poisoning, 516, 516f
- Artificial photosynthetic systems, 585
- Ascorbic acid (vitamin C)
  - deficiency of, alcohol related, 821–822
  - forms of, 822, 822f
  - function of, 439, 439t
- Asparaginase, 700
- Asparagine
  - conversion into oxaloacetate, 700
  - structure of, 32f, 33
  - synthesis of, 721, 721f
- Aspartate
  - conversion into oxaloacetate, 700
  - formation of, 720–721
  - structure of, 33, 34f, **233**
  - synthesis of, 721, 721f
- Aspartate aminotransferase, 688, 690–691
- Aspartate transcarbamoylase. *See* ATCase
- Aspartic acid, 33
- Aspartokinases, 732, 732f
- Aspartyl proteases, 261f, **261,** 262, 262f
- Aspirin
  - discovery and development of, 1043–1044
  - effect on prostaglandin H_2 synthase-1, 354f
  - as prostaglandin inhibitor, 669
- Assays
  - defined, 67
  - enzyme-linked immunosorbent (ELISA), 82–83, 83f
  - in protein purification, 67
- ATCase (aspartate transcarbamoylase)
  - active site of, 289, 289f, **289**
  - allosteric interactions in, 288–291
  - catalytic subunit, 287–288, 288f
  - concerted model, 290
  - cooperativity, 290
  - defined, 285
  - feedback inhibition, 287
  - inhibition by CTP, 758
  - reaction, 286, 286f
  - regulatory subunit, 287–288, 288f
  - separable subunits, 287–288
  - sigmoidal curve, 291, 291f
  - sigmoidal kinetics, 287, 287f
  - structure of, 288–289, 289f
  - T-to-R transition, 290–291, 290f
  - ultracentrifugation studies of, 287, 288f
- Atherosclerosis, 661, 784–785, 787–788
- Atomic motion, 971, 972f
- Atorvastatin, 1044, **1044**
- ATP (adenosine triphosphate), 108, **108,** **275,** 295t
  - as activated carrier, 435
  - binding of, 186–187, 186f, 187f
  - binding to molecular motors, 1014–1016, 1015f, 1016f
  - in Calvin cycle, 597
  - catabolism generation, 432
  - in catabolism stages, 434–435, 434f
  - as cellular energy currency, 297, 744
  - in citric acid cycle, 496–497, 508–510, 513, 514
  - defined, 424, 426
  - effect on ATCase kinetics, 292, 292f
  - electrostatic repulsion, 429
  - as energy-coupling agent, 428
  - as energy-rich molecule, 426
  - entropy increase in, 430
  - in exercise, 431, 431f, 476, 813–815, 814t
  - formation of, 424
  - generation by NADH, 436
  - hydration stabilization in, 430
  - hydrolysis of, 252
  - as immediate donor of free energy, 432
  - in insulin release, 810, 811f
  - ion gradients in, 433–434, 433f
  - in mitochondria, 550–551, 551f
  - in muscle, 813
  - in muscle contraction, 469–470
  - in nitrogen fixation, 715–718
  - phosphoryl-transfer potential, 429, 430–431, 430f
  - in random sequential reaction, 232
  - resonance structures, 429, 429f
  - structure of, 426, **427**
  - synthesis, 432–433
- ATP formation
  - 1,3-bisphosphoglycerate (1,3-BPG) and, 459–460
  - citric acid cycle and, 508
  - by cyclic photophosphorylation, 580, 580f, 581
  - enzyme-bound, 544
  - fermentation in, 462
  - from glucose to pyruvate conversion, 461–462
  - from glycolysis, 459–462
  - from phosphoryl transfer, 459–460
  - proton-motive force and, 544–545, 545f
  - pyruvate in, 460–462
- ATP hydrolysis
  - ABC transporters and, 376
  - ATP as base to promote, 277, 277f
  - in DNA replication, 832–833, 833f
  - in DNA supercoiling, 837–839, 838f
  - in driving metabolism, 427–429
  - enzyme use of, 276
  - as exergonic, 426–427
  - free energy from, 426–431
  - γ subunit in, 546
  - hydrolysis, 276
  - in kinesin motion, 1024–1026, 1025f
  - magnesium ions in, 276–277
  - myosin and, 276–278
  - myosin conformational changes, 278, 278f
  - power stroke in, 1019
  - reversible, within myosin active site, 279
  - rotational motion, 546

  in transcription, 868, 868f
  transition state formation for, 277–278
ATP synthase
  a subunit of, **543,** 544, 546–547, 546f, 578
  α subunit of, 543, **543**
  ATP-driven rotation in, 545–546, 546f
  b subunit of, **543,** 544, 578
  β subunit of, 543, **543,** 545, 546, 578
  bind-change mechanism for, 544–546, 545f
  c ring of, 545, 546–548
  of chloroplasts, 578–580
  defined, 542
  δ subunit of, **543,** 544
  discovery, 542, 542b
  efficiency of, 548, 548b
  ε subunit of, 580
  F_0 subunit of, 543, **543,** 544
  F_1 subunit of, 543, **543,** 544, 545
  G proteins and, 548–549
  γ subunit of, 543, **543,** 545, 548, 1027
  γε stalk, 544, 545
  inhibition of, 556
  nucleotide-binding sites, 545, 545f
  proton flow through, 542, 544–546
  proton motion across membrane in, 547, 547f
  proton path through membrane in, 547, 547f
  proton-conducting unit, 546–547, 546f
  regulation of, 554
  rotational catalysis in, 546, 546f
ATP synthasome, 552
ATP synthesis
  in chloroplasts, 578–580
  efficiency of, 544
  as endergonic process, 542
  mechanism, 544, 544f
  oxidative phosphorylation uncoupling from, 554–556, 555f
  proton flow around **c** ring in, 546–548
  proton gradients in, 541–549
  uncoupling electron transport from, 556–557
ATP yield, 552–553, 553t
ATP-ADP cycle, 427, 432, 432f
ATP-ADP translocase, 550–551, 551f, 557
ATPases
  AAA, 685, 1012, 1014
  domains of myosin, 276, 276f, 277, 277f
  mitochondrial. *See* ATP synthase
  Na^+-K^+, 370, 373
  P-type, 367, 370–373, 374
  SERCA, 370, 371–373, 372f, **372**
ATP-binding cassettes (ABCs). *See also* ABC transporters
  ATP binding with, 375, 375f
  defined, 374
  distance between, 375
ATP-citrate lyase, 666
ATP-grasp fold, 745
Atractyloside, 557
Atrazine, 584, **584**
Attenuation, 935–937, 936f
Autoimmune diseases, 1005, 1005f
Autoinducers, 934, 934f
Autotrophs, 566, 590
Axel, Richard, 963
Axial bonds, 320–321, 321f
Axoneme, 1023
Azothioprine, 1051, **1051**

B cells
  activation of, 993–994, 993f
  defined, 984
  memory, 1006
B vitamins. *See* Vitamins
β_2-microglobulin, 996–997
Bacillus amyloliquefaciens, 259
Bacillus anthracis, 839
Bacillus subtilis, 867
Backbone models, 61–62, 62f
Backbones
  in nucleic acids, 106–107, 106f
  in polypeptide chains, 36, 36f
Bacteria, 3, 3f. *See also Escherichia coli* (*E. coli*)
  carcinogen action on, 852, 852f
  cell membranes of, 359, 359f
  defined, 3
  flagella movement of, 1026–1029
  gene expression in, 925–938
  glyoxylate cycle in, 516–517, 517f
  nitrogen fixation in, 714–718
  photosynthetic, 584, 585t
  rapid growth, 128
  restriction enzymes in, 137
  transcription factors in, 945
  transcription in, 861–871
  translation in, 906
  upstream promoter elements in, 864
Bacterial artificial chromosomes (BACs), 147
Bacterial reaction center
  cyclic electron flow in, 572, 572f
  defined, 569–570
  electron chain in, 570, 570f
Bacteriochlorophylls, 570
Bacteriopheophytins, 570–571
Bacteriorhodopsin, 351–352, 351f
Baculovirus, 160
BAL (British anti-lewisite), 516
Ball-and-chain model, 384–385, 385f
Ball-and-stick models, 22–23, 23f, 61, 61f
Band centrifugation, 77–78, 77f
Basal transcription apparatus, 875
Base-exchange reactions, 771
Base-excision repair, 848, 849f
Base-pair substitutions, tests for, 852
Base-paired hairpins, 123
Bases/base pairs
  adenine, **4,** 5f
  anticodon, 898, 898f
  compositions, 110, 111t
  cytosine, **4,** 5f
  damage to, 846–847
  double helix, 10
  guanine, **4,** 5f
  hydrogen bonds, 5
  minor-groove side of, 830
  mutagenic, 846
  nomenclature of, 744
  non-Watson-Crick, 846, 848, 863
  nucleic acids, 109
  in nucleotides, 107–108
  purine, 107, **107**
  pyrimidine, 107, **107**
  rRNA in, 903
  rules for, 110
  sequence of, 106, 114, 124
  stacking of, 10, 10f, 111, 111f, 116
  thymine, **4,** 5f
  in translation, 896, 903
  tRNA, 896
  uracil, 107, **107**
  Watson-Crick, 110, 110f, 113, 114f, 898
Basic-leucine zipper (bZip), 945, 945f
B-cell receptor, 993, 993f
Bcr-abl gene, 417, 417f
Bcr-Abl kinase, 1048, 1054
B-DNA helix, 111
Bennett, Claude, 991
Benson, Andrew, 589
Benzoate, 697
Benzoyl CoA, 697
Berg, Howard, 1027
Beriberi, 515, 515b
Beta blockers, 1048–1049, 1050
β chains, 205, 205f
β oxidation, 813, 815, 818
β pleated sheets
  amino acid residues in, 51–52, 51t
  antiparallel, 42f
  defined, 42
  mixed, structure of, 43f
  parallel, 43f
  protein rich, 43f
  twisted, schematic, 43f
β strands
  defined, 42
  DNA recognition through, 927, 927f
  of membrane proteins, 352, 352f
  Ramachandran plot for, 42f
  structure of, 42f
β turns, 44, 51–52
β-adrenergic receptor, 399–400, 400f
β-adrenergic-receptor kinase, 404
β-chain, of hemoglobin, 191
β-globin gene
  coding sequences, 127
  structure of, 127, 127f
  transcription and processing of, 127–128, 128f
β-hemoglobin, 207–208
β-lactam ring, 239
β-oxidation pathway, 650–651
β-thalassemia, 207, 208
B-form DNA, 111–112, 111f, 112t
Bile, 788
Bile acids, 646, 646f

Bile salts
  defined, 788
  synthesis of, 788–789, 789f
Bilirubin, 737
Biliverdin, 736
Biliverdin reductase, 737
Bimolecular reactions, 226
Bimolecular sheets
  defined, 348
  formation of, 348–350
Binding-change mechanism, 544–546, 545f
Binding energy, 225, 252
Biochemistry
  defined, 1
  environmental factors influencing, 20
  evolution timeline, 3f
  as evolving science, 1–23
  genome sequencing and, 17–19
  genomic revolution and, 17–22
  unifying concepts, 1–4
Biofilms, 934–935
Biological diversity
  biochemical unity underlying, 1–4
  similarity and, 2f
Biological macromolecules, 2
Biological membranes. *See* Membranes
Biopterin, 703
Biotin
  defined, 479
  in gluconeogenesis, 479–480, 479f
Biotin carboxyl carrier protein (BCCP), 479, 479f
1,3-bisphosphoglycerate (1,3-BPG)
  as acyl phosphate, 457
  ATP formation and, 459–460
  in glycolysis, 433, 457–460, 459f
  oxidation of energy, 433
  phosphoryl-transfer potential, 430
2,3-bisphosphoglycerate (2,3-BPG)
  allosteric effector, 200–201
  binding of, 201, 201f
  defined, 200
  in oxygen affinity determination, 200–201, 201f
Bitter taste, 966, 967–968
Bjorkman, Pamela, 996
Blackburn, Elizabeth, 845
BLAST (Basic Local Alignment Search Tool) search, 177, 178f
Blood, isozymes in, 293
Blood clotting, 286, 299
  clots, 307–308
  enzymatic cascades, 303–304, 303f
  extrinsic pathway of, 304
  hemophilia and, 308
  intrinsic pathway of, 304
  regulation of, 307–308
  vitamin K in, 304, 306–307
  zymogen activation in, 303–304
Blood groups, 331–332, 331f
Blood-brain barrier, 1037
Blosum-62
  alignment of identities only versus, 176f
  defined, 174
  graphic view of, 175f
Blotting techniques, 136, 138
Boat form, 320–321, 321f
Body mass index (BMI), 802, 803f
Body weight, regulation of, 802–804
Bohr, Christian, 202
Bohr effect, 202–203, 203f
Boltzmann's constant, 222
Bombardment-mediated transformation, 163
Bonds
  axial, 320–321
  covalent, 5, 6–7
  disulfide, 36–37, 37f
  electrostatic interactions, 7
  equatorial, 320–321
  glycosidic, 322, 322f, 325, 325f, 326
  hydrogen. *See* hydrogen bonds
  hydrophobic effect and, 9, 9f
  hydrophobic interactions and, 9
  ionic interactions, 7–8, 10f
  isopeptide, 683
  length measurement, 7b
  length of, 7b
  noncovalent, 7–8
  peptide. *See* peptide bonds
  units of measure for, 7b
  van der Waals interactions, 7, 8, 8f, 10, 348, 990
Bonitus, Jacob, 515b
Bordetella pertussis, 418
Bortezomib, 687
Bovine-spongiform encephalopathy, 56
Brain
  in caloric homeostasis, 804–807
  leptin effects in, 806–807, 806f
  role in caloric homeostasis, 804–807
  sensory connections to, 962, 962f
Branch points, 880, 881f
Branched-chain amino acids, 687–689
Branched-chain ketoaciduria, 705
Branching enzyme, 631–632, 632f
Brittle bone disease, 46
Bromodomains, 953, **953**
Brown, Michael, 784
Brown adipose tissue (BAT), 555, 555f
Brown fat mitochondria, 555
Buchner, Edward, 450
Buck, Linda, 963
Buffers
  action, 15f
  pH and, 15–17
  protonation, 16f

C genes, 991–992
C_4 pathway, 599–601, 600f
C_3 plants, 601
C_4 plants, 601
CAAT box, 874, 874f
Caenorhabditis elegans, 66, 956
  genome of, 153
  RNA interference in, 163
Calcineurin, 994
Calcitriol, 795
Calcium ion channels, 382
Calcium ion pump, 372–374, 372f
Calcium ions
  calmodulin activation, 407
  coordination to multiple ligands, 406, 406f
  imaging, 406, 406f
  in phosphorylase activation, 626, 627f
  pump structure, **371**
  pumping, 372, 372f
  as second messenger, 398, **399,** 405–406
  in signal transduction, 405–407
  in vision, 972–973
Calmodulin
  defined, 626
  in mitochondrial biogenesis, 813
  in signal transduction, 407, 407f
Calmodulin-dependent protein kinases (CaM kinases), 407, 407f
Calnexin, 335
Caloric homeostasis, 802–807
  in body weight regulation, 802–804
  brain role in, 804–807
  defined, 801
  in diabetes, 807–812
  dieting and, 807
  evolution and, 802–804
  insulin in, 805–806
  leptin in, 805–807
  obesity and, 801–812
  satiation signals in, 804–805, 804f
  signaling in, 804–807
  suppressors of cytokine signaling (SOCS) in, 806–807, 807f
Calreticulin, 335
Calvin, Melvin, 589
Calvin cycle, 589–601
  ATP (adenosine triphosphate) in, 597
  C_4 pathway and, 599–601, 600f
  carbon dioxide in, 591–598, 599–601
  defined, 566, 589
  five-carbon sugar formation in, 595, 596f
  glycolate in, 593–594, 593f
  hexose sugar formation in, 594–596, 595f
  illustrated, 590f
  light regulation of, 598, 598f
  magnesium ion and, 592, 592f, 598–599
  NADPH in, 589–590, 597
  overview of, 589–590
  oxygenase reaction in, 593–594, 593f
  pentose phosphate pathway and, 610
  3-phosphoglycerate formation in, 591–592, 592f
  photorespiration in, 594, 594f
  reactions in, 590–597, 597f
  regulation of, 598–601
  ribulose 1,5-biphosphate in, 591–592, 591f, 592f, **592**
  rubisco in, 591–594, 591f, 592f, 593f, 598–599
  salvage pathway in, 593–594
  stages, 590–591, 590f

thioredoxin in, 599, 599t, **599**
cAMP. *See* Cyclic AMP
cAMP response element binding (CREB) protein, 403
Camptothecin, 839
Cancer
  aerobic glycolysis in, 474–476
  breast, 793
  citric acid cycle defects in, 513–514
  colorectal, 851, 851f, 1006
  defective DNA repair in, 850–852
  drug resistance in, 1053–1054
  fatty acid metabolism in, 667–668
  hypoxia inducible factor 1 (HIF-1) in, 513
  immune system in, 1005–1006
  in Li-Fraumeni syndrome, 851
  ovarian, 793
  pyruvate dehydrogenase kinase (PDK) in, 513
  signal transduction abnormalities in, 416–417, 416f, 417f
  skin, in xeroderma pigmentosum, 851, 851f
  tumor hypoxia in, 475–476, 476t
  tumor suppressor genes and, 851
  tumor visualization, 474, 475f
  Warburg effect in, 474
Cancer therapy, 759–760
Capillary electrophoresis, 139
Capsaicin, 977–978
Capsaicin receptor, 977–978, 977f, 978f
Captopril, 263
Carbamate
  in Calvin cycle, 592
  formation of, 204, 204f
Carbamate groups, 204f
Carbamoyl phosphate
  formation of, 693–694
  in pyrimidine synthesis, 745
  reaction, 694–695
  in urea cycle, 693–695
Carbamoyl phosphate synthetase I, 693–694
Carbamoyl phosphate synthetase II, 745
  in pyrimidine synthesis, 745
  structure of, 745, **745**
Carbanion, 606
Carbanion intermediates, 607, 607f
Carbohydrate-asparagine adduct, **57**
Carbohydrates, 315–338
  as building block, 20
  disaccharides, 323–324
  glucose generation from, 450
  isomeric forms of, 317f
  lectins and, 333–336, 334f
  monosaccharide, 316–323
  N-linked, 326, 326f
  oligosaccharides, 323, 331, 331f
  O-linked, 326
  overview of, 315–316
  polysaccharides, 324, 325f
Carbon dioxide
  binding site, 266, 266f
  in Calvin cycle, 591–598, 599–601
  defined, 203
  hydration acceleration, 264
  hydration of, 268, 268f
  oxygen release and, 202–204
  pH and, 203, 203f, 204
  transport from tissue to lungs, 204, 204f
Carbon fuels
  free energy of, 432–435, 432f
  prominent, 432, 432f
Carbon monoxide
  defined, 201
  in oxygen transport disruption, 201–202
Carbonic acid, 203
Carbonic anhydrases, 204
  bound zinc in, 265
  in carbon dioxide hydration acceleration, 264
  catalytic strategies of, 251
  defined, 264
  evolution of, 268
  mechanism of, 266, 266f
  pH effect on, 265, 265f
  proton shuttle, 267–268, 268f
  structure of, 265, **265**
  synthetic analog model system for, 266–267, 267f
Carbonium ion intermediate, 621
Carbon-nitrogen bond, 253
Carboxamides, 33
γ-carboxyglutamate, 57–58, 306–307, 306f
Carboxyhemoglobin, 201
Carboxylation, of pyruvate, 479–480, 514
Carboxyl-terminal domain (CTD)
  in coupling transcription to pre-mRNA, 884f
  phosphorylation of, 875
  serine residues of, 872
Carcinoembryonic antigen (CEA), 1006
Carcinogens, tests for, 852, 852f
Cardiotonic steroids, 373
Cargo receptor, 919
Carnitine, 649–650, 650f
Carnitine acyltransferase I, 650
Carnitine acyltransferase II, 650
Carotenoids, 581–582
Carriers, 367
Cartilage, 328–329, 329f
Caspases, 558
Cassette mutagenesis, 151, 151f
Catabolism
  ATP generation by, 432
  defined, 424, 714b
  energy from, 432–435
  free energy of, 431, 432–435, 435f
  stages of, 434–435, 434f
Catabolite activator protein (CAP), 931–932, 931f, 932f
Catabolite repression, 931–932
Catalysis
  by approximation, 252
  by ATCase, 285
  Circe effect in, 231, 231b
  covalent, 252, 253–255, 254f, 255f
  efficiency of, 230–231, 231f
  enzyme-substrate binding in, 225
  enzyme-substrate complex formation in, 222–223, 222f
  general acid-base, 252
  maximal rate, 228–229, 230
  metal ion, 252–253
  as selective binding of transition state, 241
  specificity constant, 230–231, 231f
  transition state formation and, 222
  velocity of, 231
  zinc activation of water molecule, 265–267, 266f
Catalytic antibodies, production of, 241
Catalytic strategies, 251–283
  of carbonic anhydrases, 251, 264–269
  in chymotrypsin, 253–255, 254f, 255f
  of myosins, 251–252, 275–281
  overview of, 251–252
  of restriction endonucleases, 251, 269–275
  of serine proteases, 251, 255–258
Catalytic triads
  in chymotrypsin, 255–258, 256f
  in elastase, 258, 259f
  in hydrolytic enzymes, 258–260, 258f, 259f
  serine in, 255–258
  site-directed mutagenesis and, 260, 260f
  of subtilisin, 259, 259f
  in trypsin, 258, 258f, 259f
Cataract formation, 469, 469f
Cataracts, 469, 469f
CD3, 999, 999f
CD4, 1001, 1001f, 1004–1005
CD8, 998–1000, 999f, 1004–1005
CD28, 1000
CD45, 1000
cDNA library, 149
CDP-ethanolamine, 770, **770**
Cech, Thomas, 859
Celecoxib (Celebrex), 1048, **1048**
Cell-mediated immunity, 995
Cells
  defined, 2
  hybridoma, 81–82
  motion within, 1011, 1012f
Cell-to-cell channels, 389–390, 389f, 390f
Cellular immune response, 984
Cellular respiration
  citric acid cycle in, 496–497, 496f
  enzymes in, 217
  in oxidative phosphorylation, 552–558
Cellulose, 324–325
Cell-wall biosynthesis, 1042, 1042f
Central dogma, 105
Centrifugation
  differential, 68, 68f
  sedimentation coefficients in, 76–77, 76t, 77f
  ultracentrifugation, 76–78, 77f, 287, 288f
  zonal, 77, 77f
Cephalopods, 329
Ceramidase, 775

Ceramide
  as precursor, 774
  synthesis of, 772, 772f
  in tumor growth, 774–775
Cerebroside, 346, **346,** 772
cGMP phosphodiesterase, 972
cGMP-gated ion channels, 972, 973
Chain, Ernest, 1041
Chair form, 320–321, 321f
Chanarin-Dorfman syndrome, 648
Changeux, Jean-Pierre, 198
Chaperones, 50
Charcot, Jean-Martin, 136
Charcot-Marie-Tooth disease, 1022
Chargaff, Erwin, 110
Charged tRNA, 898
Chemical shifts, 98
Chemiosmotic hypothesis, 542–543, 542f, 543f
Chemoattractants, 1028, 1029
Chemokines, 55
Chemorepellant, 1028
Chemotaxis, 1028–1029, 1029f
Cheng, Yung-Chi, 1035
Cheng-Prusoff equation, 1036
CheY, 1028
CheZ, 1028
Chirality
  of amino acids, 721, 721f
  establishment of, 717, 718f
Chitin, 329, 329f
Chlamydomonas, 887
Chloramphenicol, 914, 914t
Chlorobium thiosulfatophilum, 584
Chlorophyll *a*, 569, 569f, 582f
Chlorophyll *b*, 581–582, 582f
Chlorophylls
  defined, 568
  in photosynthesis, 568–572
  phytol, 569
  structure of, 568–569, 568f
Chlorophylls, light absorption by, 569, 569f
Chloroplasts
  ATP synthase, 578–580
  defined, 567
  electron micrograph of, 567f
  evolution of, 568
  genome of, 568
  in photosynthesis, 567–568
  reaction center, 569–570, 570f
  redox conditions, 579–580
  structure of, 567, 567f
Chlorpromazine, 1042, **1042**
Cholecystitis, 788
Cholecystokinin (CKK), 804–805
Cholera, 417–418
Choleragen, 417
Cholesterol
  "bad," 787
  carbon numbering, 790, 790f
  discovery of, 1043–1044
  excess, effects of, 785, 785f
  fatty acid chains and, 358, 358f
  formation of, 779f
  "good," 787
  levels, clinical management of, 788
  lipid rafts, 358
  as membrane fluidity regulator, 358
  in membrane lipid, 346
  as precursor for vitamin D, 794–795
  structure of, 346
Cholesterol synthesis, 776–788
  blocking of, 788
  condensation mechanism of, 778, 778f
  defined, 776, 776b
  isopentenyl pyrophosphate in, 776–777, 777f
  labeling of, 776f
  liver in, 779
  metabolism, 784
  mevalonate in, 776–777, 777f
  rate of, 779–781
  regulation of, 779–788
  site of, 782, 782f, 784
  squalene in, 777–779, 778f, 779f
  stages of, 776
  sterol regulatory element binding protein (SREBP) in, 779–781, 780f, 781f
Choline, 771
Chorismate, in amino acid synthesis, 727–729, 727f
Chromatin, 943–944, 948–954
  defined, 943
  in gene expression, 943, 948–954
  higher-order, structure of, 944, 944f
  remodeling of, 948–954, 953f
  structure of, 941, 942t
Chromatin immunoprecipitation (ChIP), 948, 948f
Chromatin-remodeling complexes, 953
Chromatography
  affinity, 70–71, 70f
  gel-filtration, 69, 69f
  high-performance liquid (HPLC), 71, 71f
  ion-exchange, 69–70, 70f
Chromogenic substrate, 254, 255f
Chromophore, 971
Chromosomes
  bacterial artificial, 147
  defined, 109
  Indian muntjac, 109, 109f
  yeast, 941, 942t
  yeast artificial, 147, 147f
Chronic myelogenous leukemia (CML), 417
Chronic obstructive pulmonary disease (COPD), 303
Chylomicron remnants, 782
Chylomicrons, 646–647, 646f, 782, 782t, 783
Chymotrypsin, 181, **182**
  active site, 255–256, 256f
  catalysis kinetics, 254–255, 255f
  catalytic mechanism, 254, 254f
  catalytic triad in, 255–258, 256f
  conformations of, 301, 301f
  covalent catalysis in, 253–255, 254f, 255f
  diisopropylphosphofluoridate (DIPF) and, 254, 254f
  elastase and, 258, 259f
  natural substrate for, **237**
  oxyanion hole, 257, 257f
  peptide hydrolysis by, 256, 257f
  S_1 pockets of, 259f
  serine residue, 253–254
  specificity of, 254f
  specificity pocket of, 258, 258f
  structure of, 255–256
  substrate preferences of, 230–231, 231t
  trypsin and, 258, 258f, 259f
Chymotrypsinogen, 255
  conformations of, 301, 301f
  defined, 299–300
  proteolytic activation of, 300, 300f
Cilia, microtubules in, 1023
Ciprofloxacin, 839
Circe effects, 231, 231b
Circular DNA, 113, 113f
Cirrhosis, 821
Cis configuration, peptide bonds, 38, 38t, 39t
Citrate
  in fatty acid metabolism, 666, 671, 672f
  in glyoxylate cycle, 517, 517f
  isomerization of, 504, 504f
  phosphofructokinase inhibition by, 472
  in polymerization, 671
  synthesis of, 513
  transport to cytoplasm, 512
Citrate synthase
  conformational changes on binding, 502, 503f
  defined, 502
  mechanism of, 502–503, 503f
Citric acid cycle, 495–518, 819
  acetyl CoA (acetyl coenzyme A) in, 495, 498–500, 501–503, 503f, 515
  aconitase in, 504, 504f
  ADP in, 406f, 506
  ATP in, 496–497, 508–510, 513, 514
  biosynthetic roles of, 514–516, 514f
  in cancer, 513–514
  carbon oxidation in, 501–510
  in cellular respiration, 496–497, 496f
  citrate isomerization in, 504, 504f
  citrate synthase in, 502–503, 503f
  crityl CoA in, 502, 503, 503f
  components of, 509t
  control sites of, 512–513, 512f
  defects in, 513–514
  entry to, 510–514
  evolution of, 516
  in exercise, 514, 514f
  glycolysis and, 497–501, 497f
  high-energy electron harvesting in, 496–497
  illustrated, 509f
  intermediates in, 514
  isocitrate dehydrogenase in, 504, 512–513, 512f
  isocitrate in, 504–505

  α-ketoglutarate in, 504–505, 505f
  α-ketoglutarate dehydrogenase complex in, 505, 512–513, 512f
  ketone bodies in, 819, 819f
  lipoamide in, 500–501, 501f
  in mitochondria, 496, 496f
  overview of, 495–497, 496f
  oxaloacetate in, 496, 502, 507–508
  pathway integration, 514, 514f
  pattern of, 496, 496f
  in photosynthesis. *See* dark reactions
  pyruvate dehydrogenase in, 511–512
  pyruvate dehydrogenase complex in, 497–501, 497f, 497t, 501f, 511–512, 511f
  pyruvate dehydrogenase phosphatase (PDP) in, 511
  rate of travel, 512, 512f
  reactions of, 508
  regulation of, 511–513
  replenishment, 514–515
  substrate channeling in, 510
  succinate dehydrogenase in, 507–508
  succinyl CoA in, 505–507, 506f
  succinyl CoA synthetase in, 506–507, 507f
  thiamine pyrophosphate (TPP) in, 499, 499f
Citrulline, 694–695
Citrullinemia, 705t
Citryl CoA (citryl coenzyme A), 502, 503, 503f
Cladribine, 760
Clamp loaders, 840, 841f
Class I MHC proteins, 996–997, 996f, 998, 1002–1003, 1003f
Class II MHC proteins, 1000–1003, 1001f
Class switching, 994–995
Clathrin, 360, 784
Clathrin-coated pit, 360
Cleavage
  DNA. *See* DNA, cleavage of
  protein, 88–89
Cleland, W. Wallace, 232
Clinical trials, 1051–1052, 1052f
Clofarabine, 760
Cloning
  of cDNA, 380
  expression, 149
  plasmid vectors in, 144–146
  in recombinant DNA technology, 147–148
Cloning vectors, 145
Closed promoter complex, 865–866, 866f
Clostridium perfringens, 464
Clostridium tetani, 1006
Clotting. *See* Blood clotting
Cluster of differentiation, 998
Coactivators, 950–951
Coat proteins (COPs), 919
Coated pits, 784
Cobalamin, 654–655, 655f. *See also* Vitamin B_{12}
Cobalt atom, 654–655
Codon-anticodon interactions, 898
Codons
  deciphered, 125, 125t
  defined, 21, 106, 123
  maximizing number of, 898
  of mitochondria, 126, 126t
  stop, 910–911, 911f
  as synonyms, 125
  in translation, 897–898, 898f
Coenzyme A, 437, **437**
Coenzyme B_{12}, 654–655, 655f, 656
Coenzyme Q, 530–531
Coenzymes
  in amino acid degradation, 682
  defined, 217
  prosthetic groups, 217
  types of, 217t
  vitamin, 438–440, 438t
Cognate DNA, 269, 272–274, 273f, **273, 274**
Cohesive ends, DNA, 144, 144f
Coiled-coil proteins, 44–45, 44f, **44,** 45f
Collagen
  amino acid sequence of, 45, 45f
  ascorbate and, 822
  in cartilage, 328
  defined, 45
  structure of, 45f
  triple helix, 45
Collagenase, 300
Color blindness, 974–975, 974f
Color vision
  in animals, 974, 974f
  cone receptors, 973–974
  defective, 974–975
  evolution of, 974, 974f
Combinational association, antibody diversity and, 992–993
Combinatorial chemistry, 185, 1045
Combinatorial control, 946
Committed step, 730
Comparative genomics, 19
Compartmentalization, 444
Compartments, drug target, 1037, 1038f
Competitive inhibition. *See also* Enzyme inhibition
  defined, 234
  dissociation constant, 235
  double-reciprocal plot, 236, 236f
  kinetics of, 235, 235f
  use of, 235
Complement cascade, 985
Complementarity-determining regions (CDRs), 988–990, 989f
Complementary DNA (cDNA)
  cloning and sequencing of, 380
  formation of, 149, 149f
  hybridization with, 135
  screening of, 149–150, 150f
Concerted model
  allosteric enzymes, 290
  defined, 198, 290
  formulation of, 211–212
  modeling oxygen binding with, 212f
  T and R states, 199f
  tetramers, 199
Condensing enzyme, 663
Cones, 970, 973–974
Conformation selection, 224
Congenital disorders of glycosylation, 332
Congenital erythropoietic porphyria, 737
Congestive heart failure, 373
Conjugation, in drug metabolism, 1037, 1038, 1039f
Connexin, 390
Connexons, 390
Conservative substitutions, 174, 175f, 176f
Controlled termination of replication, 138–139
Convergent evolution, 179, 181–182, 181f
Cooley anemia, 207
Cooperative binding
  concerted model, 198–199, 199f
  defined, 196
  oxygen delivery and, 196–197, 196f
  sequential model, 199, 199f
  substrates, 233
Copper ions, in cytochrome *c* oxidase, 536
Coproporphyrinogen III, 736, 737f
Corepressors, 931
Cori, Carl and Gerty, 637, 638
Cori cycle, 486, 486f, 493
Cori disease, 637t
Corrin ring, 654–655
Corticosteroids, 792
Cosmids, 146
Cotransporters. *See* Secondary transporters
Coulomb energy, 7
Covalent bonds
  in biological molecules, 6–7
  defined, 5
  multiple, 7
  as strongest bonds, 6–7
Covalent catalysis
  chymotrypsin as example, 253–255, 254f, 255f
  defined, 252
Covalent enzyme-bound intermediate, 458
Covalent modifications
  acetylation, 293–294, 294t
  common, 294t
  dephosphorylation, 293
  in enzyme regulation, 293–298
  irreversible, 294
  phosphorylation, 293, 294t
COX2 inhibitors, 1047–1048, 1048f
Crassulacea, 601, 601f
Crassulacean acid metabolism (CAM), 601
Creatine, **232**
Creatine kinase, 431f
Creatine phosphate, 813–814, 814t
Creutzfeldt-Jakob disease (CJD), 56
Crick, Francis, 17, 105, 109, 123, 124, 827
Critical concentration, 1018
Cro, in genetic circuit formation, 933, 933f
Cross talk, 398–399

Cross-links, 36–37, 37f
Cross-peaks, 99
CTP. *See* Cytidine triphosphate
CTP-phosphocholine cytidylyltransferase (CCT), 770
C-type lectins, 334–335, 334f
Cumulative feedback inhibition, 732
Cyanobacteria, 568, 568f
Cyanocobalamin, 655
Cyclic AMP (cAMP)
  binding to regulatory subunit, 297
  defined, 743
  effects in eukaryotic cells, 297
  in eukaryotic cells, 403
  generation by adenylate cyclase, 402–403
  in glycogen metabolism, 628–629
  phosphorylation stimulation, 403
  as second messenger, 398, **399,** 415
  in signal transduction, 402–403
  structure of, **297**
Cyclic GMP (cGMP)
  defined, 743
  as second messenger, 398, **399**
Cyclic nucleotides, 743
Cyclic photophosphorylation, 580, 580f
Cyclin-dependent protein kinases, 844
Cycloheximide, 914, 914t
Cyclooxygenase inhibitors, development of, 1047–1048
Cyclosporin, 994, **994**
Cystathionase, 726
Cystathionine, 726
Cystathionine β-synthase, 726
Cysteine
  linked, 36
  residue modification, 287f
  structure of, 32f, 33
  synthesis of, 722–723, 726
Cysteine proteases, 261–262, 261f, **261**
Cysteine thiyl radical, 754
Cystine, 36
Cytidine, 107
Cytidine diphosphodiacylglycerol (CDP-diacylglycerol), 769, **769**
Cytidine monophosphate (CMP), 769, **769**
Cytidine triphosphate (CTP), 291–292, 291f, 292f
  ATCase inhibition by, 287, 287f, 291, 758
  defined, 287
  effect on ATCase kinetics, 292, 292f
  in pyrimidine synthesis, 747–748
  T state stabilization, 291, 291f
Cytidine triphosphate synthetase, 747
Cytidylate, 108
Cytochrome
  defined, 534
  in photosynthetic reaction center, 570, 570f
Cytochrome b_5, 668, 668f
Cytochrome *bf* complex, 572, 575, 575f
Cytochrome *c*
  conformation of, 541, 541f
  evolution of, 541, 541f
  in oxidative phosphorylation, 530f, 533–534
Cytochrome *c* oxidase
  defined, 535
  electron flow in, 556
  heme, 536–537
  mechanism, 537–538, 537f
  peroxide bridge, 538, 538f
  proton transport by, 538, 538f
  in respiratory chain, 529–531, 530f, 530t, 535–538
  structure of, 536, 536f
  subunits, 536
Cytochrome P450, 1037–1038, 1038f, 1041
  defined, 790–791
  hydroxylation by, 790–791
  mechanism, 791
  protective function, 791–792
Cytoglobin, 208
Cytokines, 1002
Cytoplasm, 601, 666–667, 666f
Cytosine, **4,** 5f, 110, **110,** 850
Cytoskeleton, 1011, 1022
Cytotoxic (killer) T cells
  CD8 on, 998–1000
  defined, 984
  foreign peptide recognition, 1001, 1001f

D amino acids, 29, 29f
D genes, 992
Dalgarno, Lynn, 906
Daltons, 34, 36b
Dark reactions, 566
Darst, Seth, 862
Dawkins, Richard, 53
De novo pathways. *See also* Nucleotide synthesis
  defined, 744
  purine, 748, 748f
  pyrimidine, 744, 744f
Deamination
  in amino acid degradation, 692
  example of, 846, 846f
Death cap, 873f
Debranching enzyme, 638
Decarboxylation
  in citric acid cycle, 498
  in gluconeogenesis, 480
  in pentose phosphate pathway, 601
Decorated actin, 1018
Dedicated protein kinases, 295
Deficiency disorders, 20
Degrons, **684**
Dehydratases, 692
7-dehydrocholesterol (pro-vitamin D_3), 794–795, 794f
Deinococcus radiodurans, 828, 828f
Deleterious reactions, 456
Denisovans, 185, 185f
Density-gradient equilibrium sedimentation, 115, 115f
Deoxyadenosine, 107
5’-deoxyadenosyl radical, 655, 655f
Deoxyadenylate, 107
Deoxycytidine, 107
Deoxycytidylate, 107
Deoxyguanosine, 107
Deoxyguanosine 3′-monophosphate (3′-dGMP), 108, **108**
Deoxyguanylate, 107
Deoxyhemoglobin
  αβ dimers, 195
  carbon dioxide and, 204
  defined, 198
  quaternary structure of, 195f, **195**
Deoxymyoglobin, 192, 193, **193**
Deoxynucleoside 5′-triphosphates, 117
Deoxyribonucleic acid. *See* DNA
Deoxyribonucleoside 3′-phosphoramidites, 139, 140f, **140**
Deoxyribonucleoside triphosphate (dNTP), 829–830
Deoxyribonucleotide synthesis
  control of, 759–760
  deoxyuridylate in, 755–756, 756f
  dihydrofolate reductase in, 756
  overview of, 753
  ribonucleotide reductase in, 753–755, 755f, 756f, 759, 759f
  thymidylate, 757–758
  thymidylate in, 755–756, 756f
Deoxyribose, 4, **106,** 317, 318f, 320f
Deoxyribose phosphodiesterase, 849
Deoxyuridylate (dUMP), 755–756, 756f
Dephosphorylation, 293, 296
Deprotonation
  effect of buffer on, 268, 268f
  water, kinetics of, 267, 267f
Designer genes, 152
Desmolase, 792
Destroying angel, 873
α-Dextrinase, 450
Diabetes mellitus
  defined, 808, 808b
  glucose homeostasis in, 808
  incidence of, 808
  insulin in, 808–812
  insulin resistance and, 808–810, 810f
  ketosis in, 659–660, 660f
  obesity and, 807–812
  oxaloacetate in, 657
  type 1, 808, 812, 812f
  type 2, 808, 809–810
Diacylglycerol (DAG), 398, **399,** 647
  in lipid synthesis, 768
  in signal transduction, 405, 405f, 415
Diacylglycerol kinase, 768
Dialysis, in protein purification, 69, 69f
Diastereoisomers, 317, 317f
Diazotrophic microorganisms, 715
Dicoumarol, 307
Dictyostelium discoideum, 276, 629, 1014
Dicyclohexylcarbodiimide (DCC), 94f, 556
Diels-Alder reaction, 728
2,4-dienoyl CoA, 653, **653**

Diet
  fatty acid metabolism and, 672
  ketogenic, 660
  low-carbohydrate, 807
  low-fat, 807
  obesity and, 807
Differential centrifugation, 68, 68f
Diffusion
  facilitated, 367, 369
  lateral, 356
  lipid, 356
  membrane protein, 357
  simple, 368
  transverse, 357, 357f
Diffusion coefficient, 356, 357
Digestion
  of dietary proteins, 682, 683f
  enzymes in, 299–301, 300f
  of lipids, 645–646, 646f
  proteolytic enzymes in, 299–303
  starved-fed cycle and, 816–817
Digestive enzymes
  chymotrypsinogen, 299–301, 300f
  synthesis of, 299
Digitalis, 373
Digitoxigenin, 373, 373f
Diglyceride acyltransferase, 768
Dihydrofolate reductase, 756
Dihydrolipoyl transacetylase, 499, 500f
Dihydropteridine reductase, 703
Dihydrotestosterone (DHT), 793
Dihydroxyacetone, **316,** 317, 317f
Dihydroxyacetone phosphate (DHAP), 323
  in Calvin cycle, 595
  in glycolysis, 452f, 454–455, 455f, 456, 456f
  in lipid synthesis, 768, 768f
Diisopropylphosphofluoridate (DIPF), 237, 237f, **237,** 254, 254f
Dimerization arm, 412, 412f, 413, 413f
Dimethylallyl pyrophosphate, 777
Dimethylbenzimidazole, 655
2,4-dinitrophenol (DNP), 556–557
Dioxygenases, 704–705, 821
Diphosphatidylglycerol (cardiolipin), 770, **770**
Diphthamide, 915, 915f
Diphtheria toxin, 914–915, 915f
Dipolar ions, 29–30, 30f
Direct DNA repair, 848
Disaccharides
  defined, 323
  structure of, 323–324, 324f
  types of, 323–324, 323f, 324f
Diseases and disorders
  albinism, 705t
  alcaptonuria, 705
  alcohol-related, 821–822
  Alzheimer's disease, 56–57
  amyloidoses, 56–57
  amyotrophic lateral sclerosis (ALS), 143
  Anderson disease, 637t
  anemia, 327
  Angelman syndrome, 685
  anthrax, 839
  argininosuccinase deficiency, 697, 697f
  arsenite poisoning, 516, 516f
  atherosclerosis, 661, 784–785, 787–788
  autoimmune, 1005, 1005f
  beriberi, 515, 515b
  bovine-spongiform encephalopathy, 56
  cancer. *See* cancer
  carbamoyl phosphate synthetase deficiency, 697–698, 698f
  carnitine deficiency, 650
  cataracts, 469, 469f
  Charcot-Marie-Tooth disease, 1022
  cholera, 417–418
  citrullinemia, 705t
  congenital disorders of glycosylation, 332
  congestive heart failure, 373
  Cooley anemia, 207
  Cori disease, 637t
  coronary artery disease, 785
  Creutzfeldt-Jakob disease (CJD), 56
  deficiency disorders, 20
  diabetes. *See* diabetes mellitus
  drug-resistant, 1053–1054
  emphysema, 303
  environmental factors in, 20
  falciparum malaria, 612
  familial hypercholesterolemia, 767
  galactosemia, 468
  glucose 6-phosphate dehydrogenase deficiency, 610–612
  glycogen-storage, 637–638, 637t, 638f
  gout, 761
  Hartnup disease, 682
  heart disease, 788
  hemolytic anemia, 610–612
  hemophilia, 308
  Hers disease, 637t
  HIV infection. *See* human immunodeficiency virus (HIV) infection
  homocystinuria, 705t
  Huntington disease, 56, 683, 850
  Hurler disease, 328f, 332
  hyperammonemia, 697
  hyperlysinemia, 705t
  I-cell disease, 332
  lactose intolerance, 467–468
  Leber hereditary optic neuropathy (LHON), 557
  Lesch-Nyhan syndrome, 761–762
  Li-Fraumeni syndrome, 851
  long QT syndrome (LQTS), 388–389
  Lynch syndrome, 851, 851f
  mad cow disease, 56
  malaria, 206
  maple-syrup urine disease, 705
  McArdle disease, 637t, 638
  mercury poisoning, 515–516
  mitochondrial, 557
  mucopolysaccharidoses, 328
  multidrug resistance in, 374–376
  mutations causing, 143
  neurological, protein misfolding in, 56–57
  ornithine transcarbamoylase deficiency, 697–698, 698f
  osteoarthritis, 328
  osteogenesis imperfecta, 46
  osteomalacia, 795
  Parkinson disease, 56–57, 683, 685
  phenylketonuria, 705
  Pompe disease, 637–638, 637t, 638f
  porphyrias, 737–738
  predisposition to, 18
  prion, 56–57, 57f
  protein aggregates in, 56–57
  respiratory distress syndrome, 774
  retinitis pigmentosa, 884–885
  ricin poisoning, 915
  rickets, 795
  scrapie, 56
  scurvy, 821
  severe combined immunodeficiency (SCID), 164, 760–761
  sickle-cell anemia, 205–206, 205f, 206f
  spina bifida, 762
  Tay-Sachs disease, 774
  thalassemia, 884, 884f
  transmissible spongiform encephalopathies, 56–57
  tuberculosis, 687
  tyrosinemia, 705t
  vanishing white matter (VWM) disease, 913, 913f
  vitamin D deficiency, 795
  von Gierke disease, 637, 637t, 638
  Wernicke-Korsakoff syndrome, 821
  whooping cough, 417–418
  xeroderma pigmentosum, 851, 851t
  Zellweger syndrome, 657
Dismutation, 539b
Displacement loop (D-loop), 853
Dissociation constant (K_d), 229, 235, 1017–1018
Distal histidine, 194
Distribution, drug, 1037, 1037f, 1038f
Disulfide bonds
  formation of, 36–37, 37f
  reduction, 89, 90f
Diuron, 584, **584**
Divergent evolution, 181
DNA
  A-form, 111–112, 111f, 112f, 112t
  ancient, amplification and sequencing of, 185, 185f
  backbone of, 106f
  bases in, 4, **4,** 10, 10f, 108
  B-form, 111–112, 111f, 112t
  building blocks, 4–5
  chemically synthesized linker, 144
  circular, 113, 113f
  coding strand of, 862
  cognate, 269, 272–274, 273f, **273**
  cohesive ends, 144, 144f

DNA—(*continued*)
  complementary (cDNA), 135, 149–150, 149f, 150f, 380
  complementary single-stranded ends, 144
  covalent structure, 4, 4f
  damage to, 828, 845, 847–849
  defined, 105
  denaturation, 14–15, 14f
  directionality of, 108–109
  distortion of, 273–274, 273f
  double helix of, 5, 6, 9–10, 14–15, 105, 109–111
  double-stranded, 121, 152–153
  electroporation and, 163, 164f
  elongation, 118, 118f
  exchange between species, 184, 184f
  fingerprint, 137
  forensics and, 142, 143f
  functions of, 2
  heredity and, 5
  hybridization, 117
  information flow to RNA, 119, 119f
  ionic interactions, 10f
  length of, 108–109
  as linear polymer, 4, 105, 106
  linking number of, 834–836, 834f
  melting temperature of, 116
  methylation, 274–275
  microinjection of, 159–160, 159f
  nucleotides, 107
  promoter sites in, 122, 122b, 122f, 860
  proofreading of, 847, 848f
  properties of, 4–5
  proviral, 159
  recombinant. *See* recombinant DNA technology
  resolution by density-gradient, 115, 115f
  stability of, 142
  stem-loop structure, 113, 113f
  in storage of genetic information, 4, 5
  structure of, **108,** 111f, 113–114, 113f
  sugars, 4
  supercoiled, 113, 113f, 833–839
  synthesis of, 139–141, 140f, 141
  topoisomers of, 835–836, 836f
  transferred (T-DNA), 163
  unwinding of, 833–839, 860, 865–866, 866f
  Watson-Crick model, 111, 116
  winding, 944
  wrapping of, 943–944
  X-ray diffraction of, 109f
  Z-form, 112, 112f, 112t
DNA, cleavage of, 144, 144f
  EcoRV endonuclease, 269
  metal ion catalysis in, 271–272, 272f
  restriction enzymes and, 269–275
  stereochemistry, 271, 271f
DNA fragments
  covalent insertion of, 143–144
  insertion and replication of, 145
  insertional inactivation of, 145, 145f
  joined by DNA ligases, 144
  library of, 147–148, 148f
DNA gyrase, 839
DNA ligase, 135, 849
  defined, 832
  DNA fragments joined by, 144
  in DNA replication, 832, 844
  in forming recombinant DNA molecules, 143–144
  reaction, 832, 832f
DNA microarrays
  defined, 158
  gene-expression analysis with, 159, 159f
  using, 158–159, 158f
DNA photolyase, 848
DNA polymerase α, 843, 844, 844t
DNA polymerase β, 844t
DNA polymerase δ, 844, 844t
DNA polymerase I, 217, 829, **829,** 842, 844t
DNA polymerase II, 844t
DNA polymerase III, 831–832, 840–842, 840f, 844t
DNA polymerase switching, 843–844
DNA polymerases
  common structural features, 829
  in correcting DNA mistakes, 118
  defined, 117, 829
  DNA fragment of, 830
  error-prone, 846
  eukaryotic, 843–844, 844t
  holoenzyme, 840, 841f, 842–843
  mechanism, 829–830, 830f
  phosphodiester-bridge formation and, 117–118
  polymerization reaction catalyzed by, 117, 117f
  primer, 118, 829
  prokaryotic, 844t
  replication by, 117–119
  specificity of, 830
  structure of, 829, **829**
  template, 829
  as template-directed enzymes, 117, 829
  types of, 844t
DNA probes
  amino acid sequences in making, 91
  defined, 138
  generated from protein sequences, 147, 147f
  as primer, 140
  synthesis by automated solid-phase methods, 139–141
DNA recombination
  in cell division, 855
  defined, 852
  functions of, 853
  Holliday junctions in, 854, 854f
  illustrated, 853f
  initiation of, 853
  mechanism, 854, 854f
  overview of, 828
  RecA in, 853
  recombinases in, 854
  strand invasion in, 853, 853f
  V(D)J, 992, 992f
  VJ, 991–992, 991f
DNA recombination synapse, 854
DNA repair
  base-excision, 848, 849f
  direct, 848
  double-stand, 849
  in *E. coli*, 847–849
  enzyme complexes in, 848–849, 848f
  example of, 846, 846f
  illustrated, 828f
  ligase in, 849
  mismatch, 848, 848f
  nonhomologous end joining (NHEJ) in, 849
  overview of, 828, 845
  proofreading in, 847, 848f
  single-strand, 839–841
  thymine in, 850–851
  trinucleotide repeats and, 850
  uracil DNA glycosylase in, 850, 850f
DNA replication, 5f, 115f, 116f, 828–845
  ATP hydrolysis in, 832–833, 833f
  cell cycle and, 844, 844f
  clamp loaders in, 840, 841f
  controlled termination of, 138–139, 139f
  coordinated process in, 839–845
  cross-linking in, 847, 847f
  defective, in cancer, 850–852
  defined, 109
  DNA polymerase III in, 839, 840–842, 840f
  DNA polymerases in, 829–830, 842–843, 844t
  in *E. coli*, 115, 115f, 839–840, 842–843, 842f
  error rate, 118
  errors in, 846
  in eukaryotes, 843–844, 843f, 844f
  in eukaryotic cell cycle, 844, 844t
  helicases in, 832–833, 833f
  illustrated, 828f
  lagging strand in, 831f, 832, 840–842, 841f
  leading strand in, 831f, 832, 840–842, 841f
  licensing factors in, 843
  ligase in, 832, 832f, 844
  minor-groove interactions in, 830, 830f
  Okazaki fragments in, 831–832, 831f, 841
  origin of, 842–843, 842f, 843f
  origin of replication complexes (ORCs), 843–844
  overview of, 827
  polymerase switching in, 843–844
  by polymerases, 117–119
  prepriming complex in, 842, 842f
  primer in, 831, 831f
  processivity in, 839–840, 839f
  replication fork in, 831, 841, 841f
  replicon, 843
  RNA polymerases in, 831, 831f
  semiconservative, 115–116
  shape complementarity in, 830, 830f

  shape selectivity in, 830, 831f
  sites of, 842–844
  sliding DNA clamp in, 839–840, 839f
  specificity of, 830
  telomeres in, 844–845, 845f
  template in, 828–829
  trombone model in, 841, 841f
  tumor-suppressor genes in, 851
DNA sequencing, 89–90
  amino acid sequences and, 90f
  amplification by PCR, 141–142
  ancient DNA, 185, 185f
  by controlled termination of replication, 138–139, 139f
  decreasing costs of, 18f
  defined, 135
  as gene exploration tool, 136
  human migrations and, 19f
  methods for, 18
  next-generation (NGS), 155–156, 155f
  pyrosequencing, 155–156
  reversible terminator method, 155–156
  semiconductor, 156
DNA vectors, 143, 144–147
DnaA, 842, 842f
DnaB, 840
DNA-binding domains
  defined, 945
  in eukaryotes, 945–946
DNA-binding proteins
  basic-leucine zipper (bZip) in, 945, 945f
  binding to regulatory sites in operons, 927–932
  in eukaryotes, 945–946
  helix-turn-helix motif in, 927, 927f
  homeodomains in, 945, 945f
  in prokaryotes, 926–932
  symmetry matching in, 926–927
  transcription inhibition by, 928–931
  transcription simulation by, 931–932
  zinc-finger domains in, 945–946, 946f
DNA-binding sites, 931, 931f
  evolution of, 931
  GAL4, 948, 948f
  hypersensitive, 948
Domains
  acetyllysine-binding, 953, 953f
  defined, 3
  DNA-binding, 926–932, 945–946
  homeodomains, 945, 945f
  immunoglobulin, 986, 988–989
  nuclear-hormone receptor, 950, 950f
  protein, 48, 48f
Dopamine D2 receptors, 1042
Double helix, 5, 6
  acid-base reactions and, 14–15
  base pairs, 10
  destabilization, 14–15
  DNA molecule form, 105
  as expression of rules of chemistry, 9–10
  left-handed, 112
  minor groove, 875
  reversibly melted, 116–117
  right-handed, 112, 835
  sequence of bases, 114
  stabilization of, 109–111
  structure, 5, **5**
  topoisomerases and, 836–839, 837f, 838f
  in transcription, 875, 875f
  in transmission of hereditary information, 114–117
  unwinding of, 836
Double helix formation
  from component strands, 6
  entropy and, 13f
  heat release in, 12–13
  illustrated, 6f
  principles of, 10
Double-blind studies, 1052
Double-displacement reactions, 232, 233
Double-reciprocal plot, 229, 229f
Double-stranded DNA, 121, 152–153
Double-stranded RNA, 162, 162f
Downstream core promoter element (DPE), 874
Dreyer, William, 991
Drosophila melanogaster
  bases and genes, 66
  fluorescence micrograph of, 82f
  in genetic studies, 847
  genome of, 154, 982
  sensory bristles in, 976
  Toll receptor in, 982
Drug candidates
  absorption of, 1036–1037, 1037f
  ADME properties of, 1036–1040, 1036f
  administration routes for, 1036–1037
  characteristics of, 1034–1041
  distribution of, 1037, 1037f, 1038f
  effective concentrations of, 1035, 1035f
  excretion of, 1039–1040
  ligand binding in, 1035–1036, 1035f
  metabolism of, 1037–1039
  number of, 1045–1046
  oral bioavailability of, 1036, 1037f
  potency of, 1035–1036
  side effects of, 1035
  target compartments of, 1037, 1038f
  therapeutic index of, 1041
Drug development, 1033–1055
  animal testing in, 1044
  challenges of, 1034–1041
  clinical trials in, 1051–1052, 1052f
  combinatorial chemistry in, 1045
  genetic differences and, 1050–1051
  genomics in, 1048–1051
  high-throughput screening in, 1045
  natural products in, 1043–1044
  overview of, 1033–1034
  pharmacogenetics/pharmacogenomics in, 1050
  phases of, 1051–1052
  screening libraries in, 1044–1046, 1046f
  serendipitous observations in, 1041–1043
  7TM receptors in, 1048
  split-pool synthesis in, 1045, 1045f
  structure-based, 1046–1048, 1047f, 1048f
Drug targets, 1033, 1034f
  animal-model testing of, 1049
  compartments, 1037, 1038f
  emerging, 1050, 1050f
  identified in genomes of pathogens, 1050
  identified in human proteome, 1048–1049
Drugs
  absorption of, 1036–1037, 1037f
  administration routes for, 1036–1037
  delivery with liposomes, 349
  discovery approaches, 1033–1034, 1034f
  distribution of, 1037, 1037f, 1038f
  excretion of, 1039–1040, 1040f
  half-life of, 1040, 1040f
  metabolism of, 1037–1039
  natural products in, 1043–1044
  resistance to, 1053–1054
  response to genetic variation in, 1050–1051
  side effects of, 1035
  targets of, 1033, 1034f
  therapeutic index of, 1041
  toxicity of, 1040–1041, 1040f
Dynamic instability, 1023
Dynein
  AAA ATPases in, 1012, 1014
  ATP binding to, 1014–1016
  structure of, 1014, 1015f

E. coli. See Escherichia coli
E site, ribosomal, 905, 905f
Ear, hair cells of, 975–976, 975f, 976f
EcoRV endonuclease
  cleavage of DNA, 269, 271–272, 271f
  magnesium ion-binding site in, 271–272, 272f
  nonspecific and cognate DNA with, **274**
  recognition site structure, 272, 272f
  twofold rotational symmetry, 272, 272f
Edman degradation, 87, 87f
Effector functions, 895
EF-hand protein family, 407, 407f
Eicosanoids, 669, 669f, 670f, **670**
Elastase
  blocking of, 303
  catalytic triad in, 258, 259f
  S_1 pockets of, 259f
Electrochemical potential. *See* Membrane potential
Electron transfer
  half-reactions in, 528–529
  inhibitors, 556–557, 556f
  in oxidative phosphorylation, 526–529, 540–541, 540f
  in photosynthesis, 568–572
  rate of, 540, 540f
  uncoupling, 556–557
Electron transport chain, 497
Electron-density maps, 96–97, 96f, 97f
Electron-proton transfer reactions, 533, 533f
Electron-transferring flavoprotein (ETF), 651

Electron-transport chain
  components of, 530f, 530t
  in desaturation of fatty acids, 668–669, 668f
  electron flow through, 538, 539f
  FAD (flavin adenine dinucleotide) and, 497, 510
  $FADH_2$ (flavin adenine dinucleotide reduced) and, 526, 527
  inhibition of, 556
  iron-sulfur clusters in, 531–532, 532f
  NADH (nicotinamide adenine dinucleotide reduced) and, 497, 510, 526, 528–529
  in oxidative phosphorylation, 523, 529–541
Electron-transport potential, 526
Electrophilic catalyst, 691
Electrophoresis
  capillary, 139
  gel, 71–73, 72f, 73f, 137–138, 138f, 350–351
  SDS, 72–73, 73f
  two-dimensional, 74–75, 74f
Electroporation, 163, 164f
Electrospray ionization (ESI), 85
Electrostatic repulsion, 429
Elongation factors
  bacterial versus eukaryotic, 913
  G (EF-G), 909–910, 910f
  Ts (EF-Ts), 907–908
  Tu (EF-Tu), 907, 907f, 908
Embden-Meyerhof pathway, 450
Emphysema, 303
Enantiomers, 317, 317f
Encoding of amino acids, 124–127
Endocytosis, 784
Endoplasmic reticulum (ER), 360
  protein glycosylation in, 330, 330f
  protein synthesis in, 915–919
  protein targeting from, 915–919
  ribosome binding to, 915–919, 916f
  rough, 916
  smooth, 916
Endoplasmic reticulum (ER) stress, 811
Endoplasmic reticulum membrane, 668
Endosomes
  defined, 361, 784
  LDL release in, 786, 786f
Endosymbiosis
  mitochondria evolution from, 360
  receptor-mediated, 360–361, 360f
Endosymbiotic event, 525–526
Enediol intermediate, in glycolysis, 455, 456f
Energy. *See also* Thermodynamics
  activation, 221–222
  binding, 225, 252
  free. *See* free energy
  kinetic, 11
  light, 584–585, 970
  potential, 11
  resting human being requirement, 548b
  status, 444
  total, 11
  units of, 219b
Energy charge, 444, 444f, 482–483
  intracellular, 625
  in regulation, 554, 554f
Energy coupling agent, 428
Energy homeostasis. *See* Caloric homeostasis
Energy transfer
  from accessory pigments to reaction centers, 583, 583f
  resonance, 582–583, 582f
Enhancers
  binding sites for, 947f
  defined, 876, 946–947
  experimental demonstration of, 947f
  in gene expression, 946–947
  in transcription, 876
Enol CoA hydratase, 651–652, 652f
Enol intermediates, 502
Enol phosphate, 461
Enolase, 461
Enoyl reductase, 664
Ensemble studies, 242
Enterohepatic cycling, 1040, 1040f
Enthalpy, 11
Entropy
  change during chemical reaction, 11, 12
  double helix formation and, 13f
  overall, 12
  total, 11
Envelope form, 321, 321f
Enzymatic cascades, 303–304, 303f
Enzymatic velocity, 230
Enzyme inhibition
  affinity labels, 237–238, 237f, 238f
  competitive, 234
  by diisopropylphosphofluoridate (DIPF), 237, 237f
  illustrated, 234f
  irreversible, 234, 237–238
  mechanism-based (suicide), 238, 238f
  noncompetitive, 234–235
  reversible, 234, 234f
  by specific molecules, 234
  by transition-state analogs, 240–241, 241f
  transition-state analogs in, 234
  uncompetitive, 234
Enzyme kinetics
  for allosteric enzyme, 233, 233f
  chymotrypsin, 254–255, 255f
  of competitive inhibitors, 235, 235f
  defined, 225
  in first-order reactions, 225–226
  Lineweaver-Burk plot, 229, 229f
  maximal rate, 228–229, 230
  Michaelis constant in, 227, 228–230, 229t
  Michaelis-Menten, 229, 233, 235, 254
  Michaelis-Menten model for, 225–233
  of noncompetitive inhibitors, 235, 235f
  in pseudo-first-order reactions, 226
  in second-order reactions, 226
  sigmoidal, 287, 287f
  specificity constant, 230–231, 231f
  steady-state assumption and, 226–228, 226f
  of uncompetitive inhibitors, 235, 235f
  of water deprotonation, 267, 267f
Enzyme multiplicity, 732, 732f
Enzyme regulation
  by covalent modification, 293–298
  of protein kinase A (PKA), 297, 297f
  strategies, 285–286
Enzyme-linked immunosorbent assay (ELISA)
  defined, 82
  indirect, 82, 83f
  sandwich, 82–83, 83f
Enzyme-pyridoxamine phosphate complex (E-PMP), 690
Enzymes, 215–245
  acceleration of reactions, 221–225
  activation energy and, 221, 221f, 222
  active sites, 216, 223–224, 223f, 230, 252
  allosteric, 233, 233f, 287
  amount control, 286
  assay measurement of, 67
  ATP hydrolysis and, 276
  branching, 631–632, 632f
  catalytic activity, control of, 443–444
  catalytic efficiency, 230–231, 231f
  catalytic power, 215, 216–218
  catalytic reaction, 222–223, 222f
  in cellular respiration, 217
  classes of, 245, 245t
  cofactors for, 217, 217t
  debranching, 638f
  defined, 450b
  digestive, 299
  energy-transducing, 217–218
  ensemble studies, 242f
  homologous, 696f
  isozymes, 285–286, 292–293, 293f
  kinetic properties of, 225–233
  kinetically perfect, 456
  metabolism control of, 443
  multiple forms of, 285–286
  overview of, 215–218
  peptide-cleaving, 260–262
  in photosynthesis, 217
  P-loop structures, 280–281
  processive, 840b
  as proteins, 215
  proteolytic, 216, 286, 299–308
  of purine nucleotide synthesis, 752
  pyridoxal phosphate, 691–692, 691f, 692f
  rate enhancement by, 216, 216t
  reaction rate and, 220–221, 220f
  restriction, 137, 269–275
  reversible covalent modification, 286
  single molecule studies of, 242–243, 242f
  specificity, 215, 216–218, 217f
  spectroscopic characteristics of, 223
  substituted intermediate, 233
  substrate interaction, 223

  ternary complex of, 232
  turnover number of, 230, 230f
  urea cycle, 696–697
Enzyme-substrate binding
  in catalysis, 225
  conformation selection, 224
  induced-fit model of, 224, 224f, 252, 453
  lock-and-key model of, 224, 224f
  multiple weak attractions, 224
Enzyme-substrate complex
  defined, 222
  diffusion-controlled encounter, 231
  dissociation constant for, 229
  evidence of existence, 222–223
  fates, 226
  formation, in catalysis, 222–223, 222f
  hydrogen bonds in, 224, 224f
  reformed, 226
  structure of, 223, **223**
Epidermal growth factor (EGF)
  defined, 411
  Ras activation and, 413–414, 413f
  in signal transduction, 397, 411–414
  signaling pathway, 414, 414f
  structure of, 411–412, 411f
Epidermal growth factor receptor (EGFR)
  defined, 411–412
  dimerization of, 412, 412f, 413
  modulator structure of, 412, 412f
  overexpression of, 416–417
  phosphorylation of carboxyl-terminal tail, 413
  unactivated, structure of, 412–413, 413f
Epigenome, 942
Epimers, 317f, 318
Epinephrine
  defined, 399
  in fatty acid metabolism, 671–672
  in glycogen metabolism, 627–629, 628f, 633
  in signal transduction, 399–407
  synthesis of, 734, 734f
Epitopes, 79
Equatorial bonds, 320–321, 321f
Equilibrium constant
  defined, 219
  dissociation of water, 13–14
  between enzyme-bound reactants and products, 279
  metabolism and, 428
  rate constants, 221
  of reactions, 219–220
Equilibrium potential, 387–388, 387f
Equilibrium reaction, 220–221
Error-prone polymerases, 846
Erythropoietin (EPO), 327, 327f
Erythrose 4-phosphate, 604, 604f
Escherichia coli (*E. coli*)
  amino-terminal protein residues, 905–906
  DNA polymerase I, 829, **829**
  DNA recombination in, 853
  DNA repair in, 847–849
  DNA replication in, 115, 115f, 839–840, 842–843, 842f
  enterotoxigenic, 773
  fatty acid synthesis in, 664
  flagella, 1012, 1026–1029
  gene expression in, 925–932
  genome of, 109, 109f
  lac operon of, 929–932, 929f, 930f
  membranes, 359
  methionine repressor, 927, 927f
  promoter sequences, 865
  pyruvate dehydrogenase complex of, 497f, 497t
  RecA protein, 933
  recombinant systems and, 78
  restriction enzymes in, 269
  ribonucleotide reductase of, 753
  RNA in, 119, 119t
  RNA polymerases in, 123, 123f, 861–862, 862t
  two-dimensional electrophoresis and, 74
Essential amino acids
  defined, 682, 719
  pathway steps, 719–720, 720f
  table of, 682t
Estradiol, 793, 951
Estrogen response elements (EREs), 950
Estrogens
  defined, 789
  DNA-binding receptors and, 949
  pathways for formation of, 794f
  synthesis of, 792–793, **794**
Estrone, 793
Ethanol
  fermentation of, 462–463, 463f
  in glycolysis, 449
  metabolism, 819–822
Ethylene, 726, 726f
Eukarya, 3f
Eukarya domain, 3
Eukaryotes
  defining characteristics, 3
  DNA replication in, 843–844, 843f, 844f
  fatty acid synthesis in, 664
  glycolysis in, 451
  membranes, 359–362
  nuclear envelope, 360, 360f
  protein synthesis inhibition in, 914–915
  transcription factors in, 874
  transcription in, 871–886
  translation in, 871–872, 871f
  translation initiation in, 911–912, 912f
  tubulins in, 1023
Eukaryotic cell cycle, 844, 844f
Eukaryotic genes
  as discontinuous, 127
  mosaic nature of, 127
  organization of, 942
  quantitation and manipulation of, 157–159
Eukaryotic RNA polymerases, 860, **860**
Evolution
  of active sites, 252
  of amino acid sequences, 37
  of aminoacyl-tRNA synthetases, 902
  of blood types, 331–332
  caloric homeostasis and, 802–803
  of carbonic anhydrases, 268
  of chloroplasts, 568
  of citric acid cycle, 516
  convergent, 181–182, 181f
  of cytochrome *c*, 541, 541f
  divergent, 181
  of DNA-binding sites, 931
  of drug resistance, 1053–1054
  experimental studies of ancient DNA amplification and sequencing in, 185, 185f
  of globins, 183–184
  of glycogen metabolism, 629
  of glycogen phosphorylase, 629
  of immune system, 981–982
  of introns and exons, 128
  in the laboratory, 186, 186f
  of metabolism, 444–445
  molecular, 185–187
  obesity and, 802–803
  of oxygen, 574b
  of photoreceptors, 973–974, 974f
  of proteasomes, 686, 686f
  of proteins, 128
  of ribosomes, 903
  of signal transduction pathways, 415
  split genes in, 128
  of vitamins, 438–439
Evolutionary trees, 183–184, 183f
Excinuclease, 849
Excretion, drug, 1039–1040, 1040f
Exemestane, 793
Exercise
  ATP in, 431, 431f, 476, 813–815, 814t
  citric acid cycle in, 514, 514f
  creatine phosphate in, 813–814, 814t
  diabetes and, 811–812
  fatty acid metabolism in, 813, 813f
  fuel sources for, 813–815, 814t, 815f
  glycolysis during, 476
  mitochondrial biogenesis and, 813, 813f
  oxidative phosphorylation in, 814–815
  super compensation (carbo-loading) and, 815
Exons
  defined, 106
  in encoding protein domains, 128–129
  shuffling, 128, 128f, 129f
Expression cloning, 149
Expression vectors, 146
Extracellular signal-regulated kinases (ERKs), 414
Extrinsic pathway, clotting, 304

2-^{18}F-2-D-deoxyglucose (FDG), 474, 475f
F_{ab} fragments, 985–987
Facilitated diffusion, 367, 369
F-actin, 1017
Facultative anaerobes, 464

FAD (flavin adenine dinucleotide)
  as catalytic cofactor, 498
  defined, 435
  electron transfer potential of, 500
  structure of oxidized form, 435, 436f
  structure of reactive components of, 436, 436f
$FADH_2$ (flavin adenine dinucleotide reduced), 650–652
  defined, 436
  electron-transport chain and, 497, 510, 526, 527
  in oxidation-reduction reactions, 440, 440f
  structure of reactive components of, 436, 436f
Falciparum malaria, 612
Familial hypercholesterolemia, 767, 784–785
Faraday constant, 369, 528
Farnesyl pyrophosphate, 778, **779**
Fast switch, 464
Fasting, 817
Fat
  in adipose tissue, 815
  body, energy storage in, 645
  brown, 555
  as building block, 20
  neutral. *See* triacylglycerols
Fat cells, 645
Fatty acid chains
  melting temperature, 357, 357t
  packing of, 358, 358f
Fatty acid metabolism, 643–672
  acetyl carnitine in, 649–650
  acetyl CoA carboxylase in, 670–672, 671f
  acetyl CoA in, 512, 648–649, 650–652, 666–667, 666f
  acetyl CoA synthase in, 649
  activation in, 648–649
  acyl carrier protein (ACP) in, 661–662, 662f, **662,** 664–665, 665f
  β-oxidation pathway in, 650–651
  chylomicrons in, 646–647, 646f
  citrate in, 666, 671, 672f
  coenzyme B_{12} in, 654–655, 655f
  degradation in, 644, 644f, 651–661, 651f, 657f, 661–662
  diet and, 672
  enoyl CoA hydratase in, 651–652, 652f
  epinephrine in, 671–672
  in exercise, 813, 813f
  $FADH_2$ in, 650–652
  glucagon in, 671–672
  insulin in, 671–672
  ketone bodies in, 657–660
  lipases in, 647–648
  lipid mobilization and transport in, 645–647, 646f
  lipolysis in, 648, 648f
  malonyl CoA in, 662, 671
  methylmalonyl CoA in, 655–656, 655f, **656**
  NADH in, 650–652
  NADPH sources for, 667
  overview of, 643–645
  oxidation in, 648–649, 650–652, 651t, 656–657
  palmitoyl CoA in, 652, 652f, 653, 671
  pancreatic lipases in, 645–646, 646f
  pathway integration, 667, 667f
  peroxisomes in, 656–657, 657f
  steps in, 644–645, 644f
  triacylglycerols in, 645–648, 647f
  in tumor cells, 667–668
Fatty acid synthase
  catalytic cycle of, 664, 665f
  defined, 661, 662–663
  structure of, 664, 665f
Fatty acid synthesis
  in animals, 664–666
  in cytoplasm, 666–667, 666f
  defined, 644–645
  degradation versus, 661–662
  by fatty acid synthase, 661–668
  intermediates in, 662
  malonyl CoA in, 662
  maximal, 670–671
  reactions in, 662–664
  steps in, 644f, 662–664, 664f
Fatty acid thiokinase, 649
Fatty acids
  animals and, 660
  β oxidation of, 818
  chain length in, 343–344, 664
  cholesterol and, 358, 358f
  covalent attachment of, 643
  defined, 342
  degree of unsaturation, 343–344, 358
  derivatives, 643
  desaturation of, 668–669, 668f
  as fuel, 643, 647–652
  generation by lipolysis, 648, 648f
  in lipids, 342–344, 343f
  in muscle, 810, 815
  naturally occurring, 343t
  in obesity, 810
  odd-chain, 654, 654f
  physiological roles of, 643
  polyunsaturated, 661, 669
  saturated, 661
  structures of, 343, 343f
  unsaturated, 652–654, 652f, 653f, 661, 668–670
Fatty liver, 820
Favism, 611
F_c fragment, 985, 986
Feed forward stimulation, 472
Feedback inhibition, 285, 287, 443
  in amino acid synthesis, 731–732
  cumulative, 732
  enzyme multiplicity and, 732, 732f
  in purine synthesis control, 758–759, 758f
FeMo cofactor, 716, 716f
Fermentation
  to alcohol, 449, 462–463, 463f
  as anaerobic process, 464–465
  defined, 462
  in glycolysis, 462–464
  to lactic acid, 449, 463–464, 464f
  start and end points of, 465t
Ferredoxin, 580, 716
Ferredoxin-$NADP^+$, 576–577, 577f
Ferredoxin-thioredoxin reductase, 599
Ferritin, 736, 954–955, 954f
Ferrochelatase, 241, 736
Fetal hemoglobin, 201
Fiber types, muscle, 625, 626t
Fibrils, 325
Fibrin, 305, 305f, 308
Fibrin clots, 305–306, 305f
Fibrin monomer, 305
Fibrinogen
  conversion by thrombin, 304–306
  structure of, 304, 305f
Fibrinopeptides, 304–305
Fibrous proteins, 44–46
50S subunit, of ribosomes, 902–903, 903f, 908, 910
Fight-or-flight response, 399
Fingerprint, DNA, 137
First Law of Thermodynamics, 11
First-order reactions, 225
First-pass metabolism, 1039
Fischer, Emil, 224, 1046
Fischer projections, 22
5′ cap, 878–879, 879f
Flagella
  microtubules in, 1023
  rotary movement of, 1026–1029
  structure of, 1026, 1026f
Flagellar motor
  chemotaxis and, 1028–1029, 1029f
  components of, 1027, 1027f
  rotation of, 1026–1029
  schematic view of, 1027f
Flagellin, 1026, 1026f
Flavin adenine dinucleotide (FAD), 848
Flavin mononucleotide (FMN), 436, 436f, 532, 867, 867f
Flavoproteins, 500
Fleming, Alexander, 1041
FliG, 1027
FliM, 1027
FliN, 1027
Flip-flop, 357, 357f
Florey, Howard, 1041
Fluconazole, 1038f
Fluorescence microscopy, 82f, 84, 84f
Fluorescence recovery after photobleaching (FRAP), 356, 356f
Fluorodeoxyuridylate (F-dUMP), 757
Fluorouracil, 757
Folding funnel, 54, 54f
Foodstuffs, energy extraction from, 434–435, 434f
Formylmethionyl-tRNA, 906, 906f, 907
N-formylmethionine (fMet), 906, 907
Formyltetrahydrofolate, 749, 750, 750f

43S preinitiation complex (PIC), 912
Fourier transform, 96
Foxglove, 373, 373f
Fractional saturation, 195
François, Jacob, 928
Franklin, Rosalind, 109
Free energy
  of activation, 218
  ATP and, 426–431
  change during chemical reaction, 218–220
  defined, 12
  Gibbs, 12, 218
  in membrane transport, 369, 369f
  negative, 12
  of oxidation of single-carbon compounds, 432, 432f
  in oxidation-reduction reactions, 528
  of phosphorylated compounds, 431t
  of phosphorylation, 296
  standard change, 219–220
  thermodynamics of metabolism and, 425–426
Free radicals. *See* Reactive oxygen species
Friedreich's ataxia, 531–532
Fructofuranose, 320, 320f
Fructokinase, 465–466
Fructopyranose, 320, 320f
Fructose, 318, 318f
  activation of phosphofructokinase by, 472, 472f
  excessive consumption of, 466
  in glycolysis, 465–466, 466f
  metabolism, 466, 466f
  regulation of phosphofructokinase by, 472, 472f
  ring form, 320f
  ring structures, 320, 320f
Fructose 1,6-biphosphate
  in gluconeogenesis, 477f
  in glycolysis, 452f, 453–454, 454f
Fructose 1-phosphate, in glycolysis, 466
Fructose 2,6-biphosphate (F-2,6-BP)
  in gluconeogenesis, 483–485, 484f
  in glycolysis, 472, 472f
Fructose 6-phosphate
  in Calvin cycle, 594, 595f, 598, 598f
  in gluconeogenesis, 477f
  in glycolysis, 452f, 453, 454f, 472
  in pentose phosphate pathway, 604, 604f
  phosphorylation of, 454
Fructose biphosphatase 2 (FBPase2), 484
FtsZ, 1023
Fuel sources for, 816–817
Fumarase, 508
Fumarate
  from phenylalamine, 704, 704f
  in purine synthesis, 751
  in urea cycle, 695, 695f
4-fumarylacetoacetate, 704, **704**
Functional groups, of proteins, 28
Functional magnetic resonance imaging (fMRI), 193–194, 194f
Furan, 319
Furanose, 319, 319f, 320–321
Futile cycles, 485, 486

G elongation factor (EF-G), 909–910, 910f
G proteins
  activation of, 400–401, 401f
  ATP synthase and, 548–549
  defined, 400
  in glycogen metabolism, 627–629
  heterotrimeric, 400–402, 401f
  resetting of, 403–404, 403f
  role in signaling pathways, 401
  small, 413
  subunits, 401
G-actin, 1017
GAL4 binding sites, 948, 948f
Galactitol, 469
Galactokinase, 466
Galactolipids, 567
Galactose, 317, 318f
  defined, 466
  in glycolysis, 466–467, 466f
  missing transferase and, 468–469
  ring form, 320f
Galactose 1-phosphate uridyl transferase, 467
Galactosemia, 468
β-galactosidase, 332–333, 928, 928f
β-galactosides, 930
Galdieria sulphuraria, 184, 184f
Gangliosides, 346
  disorders of, 774
  structure of, 773, 773f
  synthesis of, 772–773, 773f
Gap junctions
  closing of, 390
  defined, 368, 389
  illustrated, 389f
  in intercellular communication, 389
  structure of, 390, 390f
Gas-phase ions, 85
Gastrointestinal peptides, 804–805, 805t
GC box, 874, 874f
GDP (guanosine diphosphate)
  in citric acid cycle, 506
  hydrolysis of, 427
  in olfaction, 964
  in signal transduction, 401–402, 413, 413f, 414
  in translation, 907–908
  in tubulin, 1023
  in vision, 972
Gel electrophoresis
  defined, 71
  isoelectric focusing and, 73, 73f
  polyacrylamide, 72, 72f
  protein separation by, 71–75
  SDS, 72–73, 73f
  SDS-PAGE, 73, 73f, 74
  two-dimensional electrophoresis and, 74–75, 74f
Gel-filtration chromatography, 69, 69f
Gellert, Martin, 836
Gene disruption, 160–162
  consequences of, 160–161, 161f
  defined, 160f
  by homologous recombination, 160, 160f
Gene expression
  analysis with microarrays, 159, 159f
  bacterial, 910
  constitutive, 925
  control of, 925–926
  as DNA information transformation, 119–124
  levels of, 157–158
  regulated, 925
  RNA in, 119–120
  transcriptional, 910, 925–926
  transcriptional, in eukaryotes, 954–957
  transcriptional, in prokaryotes, 935–937
  translational, 910, 954–956
Gene expression (eukaryotes), 941–958
  activation domains in, 945, 946
  basic-leucine zipper (bZip) in, 945, 945f
  chromatin in, 943–944, 952–953
  coactivators in, 950–951
  combinatorial control in, 946
  DNA methylation and, 949
  DNA-binding structures in, 945–946
  enhancers in, 946–947
  histones in, 943–944
  homeodomain in, 945, 945f
  hypomethylation in, 949
  in liver, 942, 942t
  mediator in, 946, 946f
  nuclear hormone receptors in, 949–951, 950f
  nucleosomes in, 943–944, 943f
  overview of, 941–943
  in pancreas, 942, 942t
  posttranscriptional, 954–957
  versus in prokaryotes, 941–943
  steroid-hormone receptors in, 951–952, 952f
  tissue-specific, 942t
  transcription factors in, 945–947
  zinc-finger domains in, 945–946, 946f
Gene expression (prokaryotes), 925–938
  attenuation in, 935–937, 936f
  autoinducers in, 934, 934f
  biofilms and, 934–935
  catabolite activator protein (CAP) in, 931–932, 931f, 932f
  catabolite repression in, 931–932
  chromatin remodeling in, 948–954, 953f
  corepressors in, 931
  DNA-binding proteins in, 926–932
  DNA-binding sites in, 931, 931f
  versus in eukaryotes, 941–943
  helix-turn-helix motif, 927, 927f
  lac operon in, 929–932, 929f, 930f
  lac repressor in, 927, 929–930, 929f, 930f
  λ repressor in, 932–933, 933f, 934f
  ligand binding in, 930
  operon model, 928–932
  overview of, 925–926

Gene expression (prokaryotes)—(*continued*)
  posttranscriptional, 935–937
  pur repressor in, 931, 931f
  quorum sensing in, 934, 934f
  regulatory sites in, 926–927
  signaling in, 933–934
  social interactions in, 934
  symmetry matching in, 927–928
  transcriptional, 925–932
"Gene guns," 163
Gene knockdown, 163
Gene knockout, 160–162, 160f, 161f
Gene therapy, 164
Gene transfer, horizontal, 184, 184f
General acid-base catalysis, 252
Genes. *See also specific genes*
  comparative analysis of, 156–157
  defined, 21
  designer, 152
  horizontal transfer, 184, 184f, 275
  number of, in human genome, 154
  regulator, 928
  reporter, 145
  RNA and, 118–119
  split, 128
  structural, 928
  synthesis by automated solid-phase methods, 139–141
  tumor suppressor, 851
Genetic code
  amino acids, 125, 125t
  aminoacyl-tRNA synthetases and, 898–902
  defined, 21, 124
  as degenerate, 125
  directionality of, 124
  features of, 125
  as nonoverlapping, 124
  universality of, 126–127
Genetically modified organisms (GMOs), 164
Genomes
  analysis of, 155–157
  of *Arabidopsis thaliana*, 154
  of *Caenorhabditis elegans*, 153
  of chloroplasts, 568
  comparative, 19, 156–157, 156f
  complete, 153f
  defined, 66
  of *Drosophila melanogaster*, 154, 982
  of *E. coli*, 109, 109f
  editing, 161–162, 162f
  of *Haemophilus influenzae*, 153, 177
  human, 154, 154f, 156, 156f
  of Indian muntjac, 109, 109f
  long interspersed elements, 154
  of mitochondria, 525–526, 525f, 526f
  next-generation sequencing, 155–156, 155f
  noncoding DNA in, 154
  number of genes in, 154
  proteome as functional representation, 66
  of puffer fish, 157, 157f
  of *Reclinomonas americana*, 526
  of *Rickettsia prowazekii*, 525–526
  of *Saccharomyces cerevisiae*, 153, 941
  sequencing of, 17–20, 152–156
  size of, 109
Genomic library
  creation of, 146–147, 147f
  defined, 147, 148
  screening of, 148, 148f
Genomic variation, 18
Genomics
  comparison, 19
  in drug discovery, 1048–1051
  revolution, 17–22
Geranyl pyrophosphate, 778, 778f
Geranyl transferase, 778, **779**
Ghrelin, 805, 805t
Gibbs free energy, 12, 218, 221
Gigaseal, 379
Gla domain, in prothrombin, 304
GlcNAcase, 327
Gleevec (imatinib mesylate), 1048, 1054
Globins. *See also* Hemoglobin; Myoglobin
  defined, 171
  evolution of, 183–184
  in human genome, 208
  sequence alignment in, 171–177
Glomeruli, 1039
Glucagon
  in essential enzyme regulation, 485
  excess in type 1 diabetes, 812
  in fatty acid metabolism, 671–672
Glucagon-like peptide 1 (GLP-1), 805, 805t
Glucocorticoids, 789
Glucogenic, amino acids, 698–703
Glucokinase, 473, 816
Gluconeogenesis, 476–487
  defined, 449, 476
  energy charge and, 478, 482–483
  fasting and, 817
  fructose 1,6-biphosphate in, 477f
  fructose 2,6-biphosphate in, 483–485, 484f
  fructose 6-phosphate in, 477f
  glucose 6-phosphate in, 477f, 481, 481f
  glucose generation and, 481
  glycolysis and, 478, 482–487
  historical perspective on, 450
  in kidney, 478
  in liver, 478, 483–485, 818
  in muscle contraction, 485–487, 486f
  overview of, 449, 476–478
  oxaloacetate in, 477f, 478–480, 819
  pathway integration, 486, 486f
  pathway of, 476–478, 477f
  phosphoenolpyruvate (PEP) in, 477f
  pyruvate conversion in, 478–480, 479f
  reactions of, 482t
  reciprocal regulation in, 482–487
  regulation of, 482–487, 812, 812f
  stoichiometry of, 481
  substrate cycles in, 485, 485f
  urea cycle in, 696, 696f
Glucose, 317–318, 318f
  in adipose tissue, 816
  as cellular fuel, 451
  characteristics of, 451
  complete oxidation of, 552–553, 553t
  conversion into pyruvate, 461–462
  in diabetes, 808
  as essential energy source, 317
  formation of, 451
  generation from glucose 6-phosphate, 481, 481f
  generation of, 450
  in glycogen, 324, 324f, 617, 618–619
  historical perspective on, 450
  homopolymers of, 324–325
  in liver-glycogen metabolism, 635–636
  metabolism, 424f
  as reducing sugar, 321–322
  ring form, 320f
  storage forms of, 324, 324f
  synthesis from noncarbohydrate precursors, 476–482
Glucose 1-phosphate, in glycogen metabolism, 619–620, 622–623
Glucose 6-phosphate (G-6P), 323
  galactose conversion into, 466–467, 467f
  in gluconeogenesis, 477f, 481, 481f, 623
  in glycolysis, 452f
  homeostasis, 816
  isomerization of, 453–454, 454f
  in liver, 622–623, 637
  metabolism of, 607–610, 618, 618f
  in pentose phosphate pathway, 602, 603f, 607–610
  transporter, 637
Glucose 6-phosphate dehydrogenase (G6PD)
  deficiency, 610–612
  in pentose phosphate pathway, 602, 610–612
  in reactive oxygen species protection, 610–612
Glucose homeostasis
  in diabetes, 807–812
  in starvation, 818–819
  starved-fed cycle in, 816–817
Glucose transporters (GLUTs), 368
  families, 473, 474t
  in glycolysis, 473–474, 474t
Glucose-alanine cycle, 692–693
Glucose-stimulated insulin secretion (GSIS), 810–812
α-1,6-glucosidase, 638
Glutamate, **233,** 700
  chirality of, 717, 718f
  formation of, 687, 692, 720–721
  in nitrogen fixation, 717–718
  oxidative deamination of, 687–689
  structure of, 33, 34f
  synthesis of, 722, 722f
Glutamate dehydrogenase, 688–689, 717, 718
Glutamate synthase, 718

Glutamine, 717–718
  degradation, 701
  in nitrogen transport, 693
  in pyrimidine synthesis, 745
  structure of, 32f, 33
  synthesis of, 722
Glutamine phosphoribosyl amidotransferase, 749
Glutamine synthetase, 717, 718, 732–733
  regulation of, 732–733, 733f
  structure of, 732–733
Glutathione, 610, 611
  conjugation reaction, 1038, **1039**
  defined, 734
  structure of, **734**
Glutathione peroxidase, 611, 735, 735f, **735**
Glutathione reductase, 611, 734
Glycan-binding proteins, 334
Glyceraldehyde, 316, **316,** 317, 317f
Glyceraldehyde 3-phosphate (GAP), 323
  in Calvin cycle, 595
  catalytic mechanism of, 458, 459f
  free-energy profiles for, 458, 458f
  in glycolysis, 452f, 454–455, 455f, 456, 456f, 457–460, 458f, 459f
  oxidation of, 432–433, 458–459, 459f
  in pentose phosphate pathway, 603, 604, 604f
  structure of, 458f, **458**
Glycerol
  generation by lipolysis, 648, 648f
  in gluconeogenesis, 476–478
  in liver, 648
  in phosphoglycerides, 344
Glycerol 3-phosphate
  in lipid synthesis, 768
  reoxidation of, 549
Glycerol 3-phosphate shuttle, 549–550, 549f
Glycerol phosphate acyltransferase, 768
Glycine
  in Calvin cycle, 593
  conversion into serine, 700
  defined, 30
  in liposomes, 349, 349f
  porphyrins from, 736–737, 737f
  structure of, 31f
  synthesis of, 722–723, 723f
Glycine synthase, 724
Glycine-glycine-glutamine (GGQ) sequence, 911
Glycoaminoglycans, 326
Glycocholate, 646f
Glycogen, 333
  branch point in, 324, 324f
  defined, 324, 617
  glucose storage in, 617, 618–619
  in glycogen metabolism, 627–629, 628f, 633
  in liver, 618, 618f, 627
  molecules, 617, 618f, 632f
  in muscle, 618, 618f
  phosphorolytic cleavage of, 620–621
  remodeling, 622, 622f
  storage efficiency, 632
  storage sites of, 618, 618f
  structure of, 618f
Glycogen lakes, 638f
Glycogen metabolism
  branching enzyme in, 631–632, 632f
  cAMP in, 628–629
  coordinate control of, 632–633, 633f
  degradation of, 618–629
  enzyme breakdown, 619–623
  epinephrine in, 627–629, 628f, 633
  evolution of, 629
  G proteins in, 627–629
  glycogen in, 627–629, 628f, 633
  glycogen phosphorylase in, 619–626
  glycogen synthase in, 630–632, 631f
  glycogen synthase kinase (GSK) in, 632, 633, 633f
  glycogenin in, 631
  α-1,6-glucosidase in, 621–622
  hormones in, 627–629, 628f
  insulin in, 635, 635f
  in liver, 635–637, 636f
  overview of, 617–619
  pathway integration, 628f
  phosphoglucomutase in, 622, 623f
  phosphoglycerate mutase in, 622
  phosphorolysis in, 619–623
  phosphorylase in, 619–627, 621f, 636, 636f
  protein kinase A (PKA) in, 629, 633, 633f
  protein phosphatase 1 (PP1) in, 629, 633–635, 634f, 636
  pyridoxal phosphate (PLP) in, 621, 621f
  regulation of, 618–619, 632–638
  regulatory cascade for, 627–629, 628f
  7TM receptors in, 627–629, 628f
  termination of, 629
  transferase in, 622, 622f
  UDP-glucose and, 630–631
Glycogen phosphorylase
  a, 623–624, 624f, 626, 627f
  allosteric regulation of, 624, 624f, 625, 625t
  amino-terminal domain of, 620–621, 620f
  b, 623, 624f, 625, 626, 627f
  carboxyl-terminal domain of, 620–621, 620f
  defined, 623
  evolution of, 629
  in glycogen metabolism, 619–626
  glycogen-binding site of, 620–621, 620f
  in liver, 623–624, 634f
  in muscle, 625, 625f
  regulation of, 623–627, 624f, 625t, 629
  structure of, 620
Glycogen synthase
  α-1,6 linkages, 631–632, 631f
  defined, 630
  in glycogen synthesis, 632
  isozymes of, 630–631
Glycogen synthase kinase (GSK)
  defined, 632
  in glycogen metabolism, 633, 633f
Glycogenin, 631
Glycogen-storage diseases, 637–638, 637t, 638f
Glycolate, 593–594, 593f
Glycolipids, 344, 345–346, 348–350
N-glycolylneuraminate, 772
Glycolysis
  aerobic, 474–476, 510
  aldehyde oxidation in, 457–458
  anaerobic mode, 510, 814
  ATP formation in, 459–462
  1,3-bisphosphoglycerate (1,3-BPG) in, 433, 457–460, 459f
  in cancer, 474–476
  citric acid cycle and, 497–501, 497f
  covalent enzyme-bound intermediate in, 458
  defined, 449
  dihydroxyacetone phosphate (DHAP) in, 452f, 454–455, 455f, 456, 456f
  as Embden-Meyerhof pathway, 450
  enediol intermediate in, 455, 456f
  energy charge and, 482–483
  as energy-conversion pathway, 451–469
  in eukaryotic cells, 451
  fermentation in, 462–464
  first stage of, 451, 451f, 452f
  fructose 1,6-bisphosphate in, 452f, 453–454, 454f
  fructose 1-phosphate in, 466
  fructose 2,6-bisphosphate in, 472, 472f
  fructose 6-phosphate in, 452f, 453, 454f, 472
  fructose in, 465–466, 466f
  galactose in, 466–467, 466f
  gluconeogenesis and, 478, 482–487
  glucose 6-phosphate (G-6P) in, 452f
  glucose transporters in, 473–474, 474t
  glucose trapping in, 451–453
  glyceraldehyde 3-phosphate (GAP) in, 452f, 454–455, 455f, 456, 456f, 457–460, 458f, 459f
  hexokinase in, 451–453, 453f, 470–471, 473
  historical perspective on, 450
  insulin in, 816
  isomerization of carbon sugars in, 453–454, 454f
  in liver, 472–473, 483–485
  in muscle, 469–471, 471f
  NAD^+ in, 458, 459f, 462–464, 465
  NADH in, 458, 462–464
  overview of, 449, 451
  pathway integration, 486, 486f
  pentose phosphate pathway and, 602–610
  phosphofructokinase (PFK) in, 469–470, 470f, 472, 472f
  phosphoglycerate kinase in, 459
  phosphoglycerate mutase in, 460
  2-phosphoglycerate in, 460–461, 460f
  3-phosphoglycerate in, 460–461, 460f
  pyruvate in, 452f, 459–464

- Glycolysis—(*continued*)
  - pyruvate kinase in, 471, 473, 473f
  - reactions of, 462t–463t
  - reciprocal regulation in, 482–487
  - regulation of, 812, 812f
  - second stage of, 451, 452f, 457f
  - stages of, 451, 451f, 452f, 457f
  - thioester intermediate in, 457–458, 458f
  - triose phosphate isomerase (TPI) in, 455–456, 455f, 456f
- Glycolytic intermediates, 610
- Glycolytic pathway
  - glucokinase in, 473
  - hexokinase in, 470–471, 473
  - phosphofructokinase (PFK) in, 469–470, 470f, 472, 472f
  - pyruvate kinase in, 471, 473, 473f
  - role of, 469
- Glycopeptide transpeptidase, 239
- Glycophorin, 336, 355, 355f
- Glycoproteins, 325–329
- Glycosaminoglycans, 327–328, 328f
- α-1,6-glucosidase, in glycogen metabolism, 621–622
- Glycosidic bonds, 322, 322f, 325, 325f, 326f
- Glycosylation, 330, 331, 331f
  - congenital disorders of, 332
  - errors in, 332
  - in nutrient sensing, 327, 327f
- Glycosyltransferases, 331, 331f
- Glyoxylate, 517, 517f, 593
- Glyoxylate cycle, 516–517, 517f
- Glyoxysomes, 517
- GMP (guanosine monophosphate)
  - in purine synthesis, 751–752, 751f
  - synthesis of, 758–759, 758f
- GMP synthetase, 752
- Goldstein, Joseph, 784
- Golgi complex, 330, 330f, 332, 918
- Gout, 761
- G-protein receptor kinase 2 (GRK2), 404
- G-protein-coupled receptors (GPCRs), 402
- Gradient centrifugation, 77–78, 77f
- Granzymes, 1000
- Grb2, 413, 413f
- Green fluorescent protein (GFP), 58–59, 58f, 84f
- Greider, Carol, 1, 845
- Group-specific reagents, 237
- Group-transfer reactions, 441, 441f
- GTP (guanosine triphosphate)
  - as energy source, 743
  - hydrolysis, 403–404
  - in olfaction, 964
  - in resetting G proteins, 403–404
  - in signal transduction, 401–402
  - in translation, 907–908
  - in tubulin, 1023
  - in vision, 972
- GTPases
  - defined, 414
  - Ras family of, 414, 414t
  - in signal transduction, 414, 414t
- Guanidinium chloride, 49
- Guanine, **4,** 5f, 110, **110**
- Guanine-nucleotide-exchange factor (GEF), 413
- Guanosine, 107
- Guanylate, 108, 752
- Guanylate cyclase, 973
- Guide strand, 162–163
- Gustation. *See* Taste
- Gustducin, 967, 967f

- H zone, 1020, 1020f
- *Haemophilus influenzae*, 153, 177
- Hair cells, 975–976, 975f, 976f
- Hairpin turns. *See* Reverse turns
- Half-life
  - of drugs, 1040, 1040f
  - of proteins, 683, 684, 685f
- Half-reactions
  - defined, 527
  - in electron transport, 528–529
- Haptenic determinants, 994
- Haptens, 994
- Hartnup disease, 682
- Hatch, Marshall Davidson, 599
- Haworth projections, 319
- Hearing, 975–977
  - hair cells in, 975–976, 975f, 976f
  - ion channels in, 976–977
  - mechanism of action, 975
- Heart disease, 788
- Heat
  - defined, 11
  - oxidative phosphorylation uncoupling in generation of, 554
  - release in double helix formation, 12–13
  - thermodynamics and, 11
- Heat shock protein 70 (Hsp70), 179, **179**
- Heat-shock response element, 875–876, 876f
- Heat-shock transcription factor (HSTF), 875–876
- Heavy (H) chains
  - in antibody diversity, 992–993
  - antigen-binding and, 991–992
  - defined, 985–986
  - in immunoglobulin classes, 987t
- Heavy meromyosin (HMM), 1013
- Heinz bodies, 612, 612f
- Helicases
  - defined, 832
  - in DNA replication, 832–833
  - mechanism, 833, 833f
  - structure of, 832, **832**
  - symmetry, 833, 833f
- Helix-turn-helix motif, 48, 48f, 927, 927f
- Helper T cells
  - action, 1002, 1002f
  - in cytokine secretion, 1002
  - defined, 984, 1000
  - in HIV infection, 1003–1004
  - peptide recognition, 1001, 1001f
  - T-cell receptors (TCRs) and, 1000–1002
- Hemagglutinin, 335–336, 335f
- Heme
  - biosynthetic pathway, 737f
  - in cytochrome *bf* complex, 575
  - in cytochrome *c* oxidase, 536–537
  - defined, 46
  - intrinsic reactivity of, 194–195
  - labeling of, 735b, 736, 736f
  - in oxygen-binding, 192, 193–194, 194f
- Heme prosthetic group, 534
- Hemichannels, 390
- Hemoglobin, 191–212
  - $\alpha_2\beta_2$ tetramer of, 49f
  - alignment comparison, 173–174, 173f
  - alignment with gap insertion, 173, 173f
  - αβ dimers, 195
  - α chains, 195, 207–208
  - amino acid sequences of, 171–172, 171f, 172f
  - β chains, 191, 195, 205, 205f
  - conformational changes in, 200, 200f
  - cooperativity, 198–199
  - defined, 171, 191
  - fetal, 201
  - globin curve, 195
  - Hill plot, 210f
  - myoglobin comparison, 191–192
  - oxygen delivery by, 196, 196f
  - oxygen-binding, 192–202, 196f
  - oxygen-binding curve, 196, 196f
  - polypeptide chains, 192, 195
  - purified, 200, 200f
  - quaternary structure of, 195, **195,** 197–198, 198f
  - R state, 198, 199f, 200, 211–212
  - sequence alignment of, 176, 176f
  - sickle-cell, 205, 205f
  - space-filling model, 195f
  - T state, 198, 199f, 200, 203, 211–212
  - tertiary structure of, 178–179
  - thalassemia and, 207
- Hemoglobin A (HbA), 195
- Hemoglobin H (HbH), 207
- Hemoglobin S (HbS), 205–206, 206f
- Hemolytic anemia, G6PD deficiency and, 610–612
- Hemophilia, 308
- Hemorrhagic disposition, 307b
- Hemostasis, 303
- Henderson-Hasselbalch equation, 16
- Heparin, 307
- Heptoses, 316
- HER2 receptor, 413, 417
- Herbicides, photosynthesis and, 584
- Hereditary nonpolyposis colorectal cancer (HNPCC), 851, 851f
- Hers disease, 637t
- Heterodimeric sweet receptor, 968–969, 968f
- Heterotrimeric G proteins, 400–402, 401f
- Heterotrophs, 566, 590
- Heterotropic effects, 292
- Hexokinase
  - defined, 451
  - in glycolysis, 451–453, 453f

  inhibition of, 471
  liver and glycolysis and, 473
  metal ion requirement, 453
  muscle and glycolysis and, 470–471
Hexose monophosphate pathway. *See* Pentose phosphate pathway
Hexose monophosphate pool, 595
Hexose sugars, formation in Calvin cycle, 594–596, 595f
Hexoses, 316, 453, 596f
High-density lipoproteins (HDLs), 782, 782t
  defined, 783
  properties of, 787
  protective effects and, 787–788
  reverse cholesterol transport in, 787
High-performance liquid chromatography (HPLC), 71, 71f
High-throughput screening, 1045
Hill, Archibald, 210
Hill coefficient, 210, 211f
Hill plot, 210–211, 210f, 211f
Histamine, 734, 734f
Histidine
  degradation, 700, 700f
  distal, 194
  ionization of, 33
  proton shuttle, 268, 268f
  proximal, 193, 200
  structure of, 33, 33f
Histidine operon, 936–937, 936f
Histone acetyltransferases (HATs), 952, **952**
Histone code, 954
Histone deacetylases, 953–954
Histone octamer, 943–944
Histones
  acetylation of, 952–953
  amino-terminal tails of, 952–953
  defined, 943
  in gene regulation, 943–944, 944f, 952–954, 954t
  homologous, 944, 944f
  modifications, 954, 954t
HIV. *See* Human immunodeficiency virus (HIV) infection
HMG-CoA reductase, 1044
HMG-CoA reductase inhibitor, 1044–1045
Hodgkin, Alan, 378
Hoffman, Felix, 1043
Hogness box, 122
Holliday, Robin, 854f
Holliday junctions, 854, 854f
Homeodomain, 945, 945f
Homeostasis, 442, 443f. *See also* Caloric homeostasis; Glucose homeostasis
Homocysteine
  in activated methyl cycle, 725, 725f
  in amino acid synthesis, 725–726
  in cysteine synthesis, 726
  levels, reducing, 726
  in vascular disease, 726
Homocysteine methyltransferase, 725
Homocystinuria, 705t
Homogenization, in protein purification, 67
Homogentisate, 704, **704,** 705
Homogentisate oxidase, 704
Homologies
  databases for, 177, 178f
  defined, 170
  detection of, 171–177
  evolutionary trees and, 183–184, 183f
  repeated motifs and, 180–181
  sequence alignment in, 173–177, 173f, 179–180
  shuffling in, 173–174, 173f
  statistical analysis in, 173–174
  substitution matrices in, 174–177
  three-dimensional structures and, 179–180
Homologous recombination, 160, 160f
Homologous sequences, 177
Homologs, 170, 171f
Homolytic cleavage reaction, 655, 655f
Homopolymers, 324
Homotropic effects, 290
Hood, Leroy, 991
Horizontal gene transfer, 184, 184f, 275
Hormone receptors, 950–952, 951f
Hormone-receptor complex, 402
Hormones
  in acetyl CoA carboxylase regulation, 671–672
  eicosanoid, 669, 669f, 670f, **670**
  in fatty acid metabolism, 670–672
  in gene expression, 950–952
  in glucose metabolism, 485
  in glycogen metabolism, 627–629, 628f
  local, 669
  metabolic functions of, 443–444
  in gluconeogenesis, 485
  in signal transduction, 400–402
  steroid, 789–791, 789f
Human genome, 154, 154f, 156, 156f
Human immunodeficiency virus (HIV) infection
  classes of, 1003
  drug development effects, 1047, 1047f
  helper T cells in, 1003–1004
  host cell for, 1003
  immune system and, 1003–1004
  protease, 262, 263f
  protease inhibitors, 263, 263f, 792, 1046, 1046f, 1053
  receptor, 1003f
  schematic representation, 1003f
  vaccine for, 1007
Humoral immune response, 984
Huntington disease, 56, 683, 850
Hurler disease, 328f, 332
Huxley, Andrew, 378
Hybridoma cells, 81–82
Hydrogen bonds, 8f
  acceptors, 15
  in base pairs, 5
  breaking with membranes, 354
  in enzyme-substrate complex, 224, 224f
  interaction, 8
  as noncovalent bond, 7
  strength of, 8
  water, 9, 10
Hydrogen ions
  in acid-base reactions, 13
  concentration in solutions, 13
  oxygen release and, 202–204
Hydrogen-bond donors, 8
Hydrolases, 245t
Hydrolysis. *See also* ATP hydrolysis
  GTP (guanosine triphosphate), 403–404
  of lipids, 647–648
  in metabolism, 427–429, 441–442, 442f
  of peptide bonds, 253, 256, 257f
  of phosphodiester bond, 269–270, 269f
Hydronium ions, 13
Hydropathy plots, 355, 356f
Hydrophilic amino acids, 33, 47
Hydrophilic moiety, 347, 347f
Hydrophobic amino acids, 30–32
  structures of, 31f
  types of, 30–32
Hydrophobic effect, 9, 9f, 10, 32
Hydrophobic interactions
  defined, 9
  as noncovalent bond, 7
Hydrophobic moiety, 347, 347f
3-hydroxy-3-methylglutaryl CoA, 702
3-hydroxyanthranilate, 704, **704**
25-hydroxycholesterol, 781
2-hydroxyglutarate, 513
Hydroxyl groups, 32–33, 57
Hydroxylation
  cytochrome P450 in, 791–792, 791f
  steroid hormones, 790–791
3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase), 776–777, 777f, 779, 781f, 788, 788f
P-hydroxyphenylpyruvate, 704, **704,** 728
P-hydroxyphenylpyruvate hydroxylase, 704
Hydroxyproline, **57**
Hyperammonemia, 697
Hyperbaric oxygen therapy, 202
Hyperlysinemia, 705t
Hypersensitive sites, 948
Hypervariable loops, 988–990, 989f
Hypochromism, 116, 116f
Hypomethylation, 949
Hypoxanthine, 761
Hypoxanthine-guanine phosphoribosyltransferase (HGPRT), 752, 761
Hypoxia
  defined, 475
  tumor, 475–476, 476t
Hypoxia-inducible factor 1 (HIF-1), 476, 476f, 513

I band, 1020, 1020f
Ibuprofen, 1038, 1038f
Ice, 9, 9f
I-cell disease, 332, 332f
Imatinib mesylate (Gleevec), 1048, 1054

Immune response
  basis of, 985
  cellular, 984
  development of, critics of, 985
  humoral, 984
Immune system, 981–1009
  adaptive, 984–985
  in autoimmune diseases, 1005, 1005f
  B cells in, 984, 993–994, 1006
  in cancer, 1005–1006
  in disease, 1004–1007
  evolution of, 981–982
  innate, 982–983
  MHC proteins in, 995–1004
  overview of, 981–982
  T cells. *See* T cells
  T-cell receptors. *See* T-cell receptors (TCRs)
Immunity, cell-mediated, 995
Immunizations, 1006–1007
Immunoelectron microscopy, 84, 84f
Immunoglobulin A (IgA), 987, 987t
Immunoglobulin D (IgD), 987, 987t
Immunoglobulin domains. *See also* Immunoglobulin fold
  constant, 988, 988f
  defined, 986
  variable, 988, 988f, 989f
Immunoglobulin E (IgE), 987–988, 987t
Immunoglobulin fold
  complementarity-determining regions (CDRs), 988–990, 989f
  defined, 988
  elements of, 988–989
  structure of, **989**
Immunoglobulin G (IgG)
  antigen cross-linking, 986, 986f
  cleavage of, 986, 986f
  defined, 985
  properties of, 987t
  structure of, 985–986, 986f
Immunoglobulin M (IgM)
  defined, 987
  properties of, 987t
Immunoglobulins. *See also* Antibodies
  classes of, 986–988, 987f, 987t
  defined, 79, 984
  heavy chains, 985–987, 987t, 991–993
  light chains, 985–987, 987t, 991–993, 991f, 992f
  production of, 984, 984f
  segmental flexibility of, 987, 987f
  sequence diversity in, 988f
  structure of, 79f
Immunologic techniques
  enzyme-linked immunosorbent assay (ELISA), 82–83, 83f
  fluorescence microscopy, 84, 84f
  hybridoma cells in, 81–82
  monoclonal antibodies in, 80–82, 80f, 81f
  polyclonal antibodies in, 80, 80f
  in protein studies, 79–84
  western blotting, 83–84, 83f
Immunological memory, 1006
Immunoreceptor tyrosine-based activation motif (ITAM), 993, 993f
Immunosuppression, 994
Immunotoxins, 152
Indian muntjac, 109, 109f
Indinavir (Crixivan), 263, 263f
Indinavir complex, 263–264, 263f
Indirect ELISA, 82, 83f
Induced pluripotent stem (iPS) cells, 947, 947f
Induced-fit model, 224, 224f, 252
Influenza virus, 335–336, 335f
Inhibition
  competitive, 235, 235f, 236f
  noncompetitive, 235–236, 235f, 236f
  reversible, 235–236, 235f, 236f
  uncompetitive, 235, 235f, 236f
Inhibition constant (K_i), 1036
Inhibitory factor 1 (IF1), 554
Innate immune system, 982–983
Inosinate, 759
Inosine, 760, 761, 898
Inositol 1,4,5-trisphosphate (IP_3), 398, **399,** 404, 405f, 415
Insertional inactivation, 145, 145f
Insulin
  amino acid sequence of, 37f
  binding of, 408–409, 409f
  defined, 27
  in diabetes, 808–812
  in essential enzyme regulation, 485
  in fatty acid metabolism, 671–672
  in glycogen metabolism, 635, 635f
  in glycolysis, 816
  MALDI-TOF mass spectrum of, 86f
  in obesity, 805–806
  pancreatic failure and, 810–812
  in protein synthesis, 816
  release of, 810, 811f
  secretion of, 805
  in signal transduction, 407–411
  signal transduction, 808–809, 808f
Insulin receptor
  activation of, 408–409, 409f
  binding sites for, 808
  defined, 408
  structure of, 408, 408f
  subunits, 408, 408f
Insulin resistance
  biochemical basis of, 808–809
  in muscle, 810–812
  in pancreatic failure, 810–812, 811f
Insulin signaling
  illustrated, 409f
  lipid kinase in, 410, 410f
  pathway, 411, 411f
  termination of, 411
Insulin-receptor substrates (IRS), 409–411, 410f, 635
Integral membrane proteins, 351, 351f
Interatomic distances, 7b
Intercalation, 870
Intermediary metabolism. *See* Metabolism
Internal guide sequence (IGS), 887
Internal ribosome entry sites (IRES), 912
Intracellular pathogen, 995, 995f
Intramolecular hemiacetal, 318
Intramolecular hemiketal, 318–319
Intramolecular rearrangements, 654, 655f
Intrinsic pathway, clotting, 304
Intrinsically unstructured proteins, 55–56
Introns
  defined, 106
  detection of, 127, 127f
  self-splicing, 887, 887f
Inverted repeats, 272, 272f
Ion channels, 378–389
  acetylcholine receptor, 385–387, 386f, 387f
  action potentials and, 378, 378f, 387–388
  ball-and-chain model for inactivation, 384–385, 385f
  calcium, 382
  cell-to-cell, 389–390, 389f, 390f
  cGMP-gated, 972, 973
  defined, 368, 378
  disruption of, 388–389
  energetic basis of selectivity, 382, 382f
  equilibrium potential and, 387–388, 387f
  gap junction, 389–390, 389f, 390f
  in hearing, 976–977
  inactivation of, 384–385, 385f
  ligand-gated, 386
  membrane permeability and, 390–391
  nerve impulses and, 378
  patch-clamp technique, 379, 379f
  potassium, 378, 379–385, 381f, **381,** 384f, 385f
  selectivity filter, 381, 381f, 382, 382f
  sequence relations of, 380, 380f
  shaker, 380
  sodium, 378, 382–385, 382f, 383f
  in taste, 969, 969f
  transient receptor potential (TRP), 977
  voltage-gated, 383–384, 384f
Ion gradients, 433–434, 433f
Ion semiconductor sequencing, 156
Ion-exchange chromatography
  defined, 69–70
  illustrated, 70f
Ionic interactions, 7
IRE-binding protein (IRP), 955–956, 956f
Iron
  metabolism of, 954–956
  oxygen-binding to, 192–195, 194f
  structural transition in subunit, 200
Iron center, of ribonucleotide reductase, 753, 753f
Iron-response element (IRE), 955, 956
Iron-sulfur clusters, 531–532, 532f
Iron-sulfur proteins
  defined, 531
  in nitrogen fixation, 715–716, 716f
  in oxidative phosphorylation, 531–532
  structure of, **715**

Irreversible inhibition. *See also* Enzyme inhibition
  defined, 234
  in mapping active site, 237–238
  mechanism-based (suicide), 238, 238f
Isocitrate
  in citric acid cycle, 504–505
  in glyoxylate cycle, 517, 517f
Isocitrate dehydrogenase
  in citric acid cycle, 504, 512–513, 512f
  mutations in, 513
Isocitrate lyase, 517
Isoelectric focusing, 73, 73f
Isoform, 292
Isoleucine, 31f, 32
Isomerases, 245t
Isomerization reactions, in metabolism, 441, 441f
Isomers
  constitutional, 317, 317f
  diastereoisomers, 317, 317f
  stereoisomers, 317, 317f
Isopentenyl pyrophosphate, 776–777, 777f, **777**
Isopeptide bonds, 683
Isopropylthiogalactoside (IPTG), 930
Isotope labeling studies, 510, 736
Isovaleryl CoA, 702
Isovaleryl CoA dehydrogenase, 702
Isozymes
  in blood, 293
  defined, 292
  distinguishing, 292
  expression of, 286
  as homologous, 285
  of lactate dehydrogenase (LDH), 292, 293f
  in varying regulation, 285–286

J genes, 991–992
Jagendorf's demonstration, 578, 578f
Jenner, Edward, 1006

K^+ ions. *See* Potassium ions
k_{cat} (rate constant), 230
k_{cat}/K_M (specificity constant), 230–231, 231f
K_d (dissociation constant), 229, 235, 1017–1018
Kendrew, John, 192
Ketimine, **690**
α-ketoacids, 701–702, 720–721
β-ketoacyl synthase, 663
α-ketobutyrate, 701
Ketogenic diets, 660
Ketoglutarate, **233**
α-ketoglutarate
  in amino acid degradation, 698, 700–701, 700f
  in amino acid synthesis, 717–718, 720
  in citric acid cycle, 504–505, 505f
  in double-displacement reactions, 233
  in glutamate synthesis, 717, 718
  in nitrogen fixation, 717
  oxidative decarboxylation of, 505
α-ketoglutarate dehydrogenase complex
  in citric acid cycle, 512–513, 512f
  defined, 505
α-ketoisocaproate, 701
Ketone bodies
  acetyl CoA formation from, 657–658
  in citric acid cycle, 819, 819f
  defined, 657
  in diabetes, 659–660, 660f
  in fatty acid metabolism, 657–660
  formation of, 657–658, 658f
  as fuel source, 658–660
  high levels of, 659–660
  in liver, 659, 659f, 819, 819f
  as water soluble, 659
Ketoses
  aldoses conversion to, 454
  defined, 316
3-ketosphinganine, 772
β-ketothiolase, 651
Keyhole limpet hemocyanin (KLH), 994
Khorana, Har Gobind, 124
K_i (inhibition constant), 1036
Kidney
  drug excretion in, 1039
  gluconeogenesis in, 478
Killed (inactivated) vaccines, 1006
Killer T cells. *See* Cytotoxic (killer) T cells
Kilobase (kb), 120b
Kilocalories (kcal), 219
Kilodaltons (kDa), 34, 36b
Kilojoules (kJ), 219
Kinase fold, 298
Kinases, 453, 453f. *See also specific kinases*
Kinesin
  ADP in, 1025, 1025f
  ATP binding to, 1014–1016, 1016f
  ATP hydrolysis in, 1024–1026, 1025f
  defined, 1012
  gene mutations, 1022
  monitoring movements mediated by, 1024f
  motion processivity, 1024–1026
  movement along microtubules, 1022, 1024–1026, 1025f
  neck linker, 1016, 1016f
  P-loop NTPase, 1012
  relay helix, 1016, 1016f
  structure of, 1014, 1014f
  walk, 1025
Kinetic energy, 11
Kinetics
  defined, 225
  enzyme. *See* enzyme kinetics
  as study of reaction rates, 225–226
  of water deprotonation, 267, 267f
Klenow fragment, 829
Klug, Aaron, 943
K_M. *See* Michaelis constant
Kornberg, Roger, 862
Krebs, Hans, 434, 516b
Krebs cycle. *See* Citric acid cycle
Kringle domain, 304
Kühne, Friedrich Wilhelm, 450b

L amino acids, 29, 29f
Laborit, Henri, 1042
Lac operon, 929–932, 929f, 930f
Lac regulatory site, 926f
Lac repressor
  binding of, 929
  components of, 929
  defined, 927
  effects of IPTG on, 930f
  inhibition relief, 931–932
  structure of, 929, 929f, 930
Lac repressor-DNA complex, 926–927, 926f
Lactase, 324
  deficiency of, 467–468
  defined, 450
Lactate, **232**
  in Cori cycle, 486, 486f
  formation from pyruvate, 463–464
  formation of, 485–486, 486f
  in gluconeogenesis, 476
  in glycolysis, 449
  in muscle contraction, 485–486, 486f
Lactate dehydrogenase (LDH), 292, 293f, 487
Lactic acid fermentation, 449, 463–464
Lactobacillus, 468f
β-lactoglobulin, 86f
Lactonases, 602
Lactose
  defined, 323–324
  structure of, 323–324, 324f
Lactose intolerance, 467–468
Lactose operon. *See Lac* operon; *Lac* repressor
Lactose permease, 377, 377f, **377**
Lagging strand
  defined, 832
  synthesis of, 831f, 840–842
  template, 841
γ chains, 987, 987f
λ phages
  alternative infection modes for, 146, 146f
  for DNA cloning, 144–147
  mutant, as cloning vector, 146, 147f
λ repressor
  defined, 933
  in gene expression, 932–933
  in genetic circuit formation, 933, 933f
  structure of, 933f
  synthesis control, 934, 934f
Lamprey, 183, 183f
Lanosterol, 778, **779,** 781
Lansoprazole, 1049, **1049**
Lateral diffusion, 356, 357, 357f
Lavoisier, Antoine, 714
Laws of thermodynamics, 10–12
Leader peptide sequences, 936, 936f
Leader sequences, 935–936, 935f

Leading strand
  defined, 832
  synthesis of, 831f, 840–842
Leber hereditary optic neuropathy (LHON), 557
Lectins
  as carbohydrate-binding proteins, 334–336
  classes, 334–335
  C-type, 334–335, 334f
  defined, 334
  in interactions between cells, 334
Leder, Philip, 991
Leghemoglobin
  sequence alignment of, 176, 176f
  tertiary structure of, 178–179
Leptin
  in caloric homeostasis, 805–807
  defined, 805
  effects in brain, 806–807, 806f
  obesity and, 806–807, 806f
  secretion of, 805, 806
Leptin resistance, 806–807, 807f
Lesch-Nyhan syndrome, 761–762
Leucine, 31f, 32
Leucine-rich repeats (LRRs), 982
Leukemia, 474–475
Leukocyte antigen A2 (HLA-A2), 996–997, 997f
Lever arm, 1015, 1015f, 1022, 1022f
Levinthal, Cyrus, 53
Levinthal's paradox, 53
Licensing factors, 843
Li-Fraumeni syndrome, 851
Ligand binding
  concerted model, 198–199, 199f
  in gene expression, 930
  sequential model, 199, 199f
Ligand-gated channels, 386
Ligands, 950
Ligases, 245t
Ligation reactions, in metabolism, 440–441
Light
  absorption, 971–972
  electromagnetic spectrum of, 970, 970f
  in lowering calcium level, 972–973
Light (L) chains, 985–987, 987t, 991–992, 991f, 992f
Light absorption, 569, 569f
Light energy, 584–585, 970
Light meromyosin (LMM), 1013
Light reactions. *See also* Photosynthesis
  in Calvin cycle, 598, 598f
  defined, 566
  herbicides and, 584
  illustrated, 566f
  stoichiometry for, 581
Light-harvesting complexes, 583
Light-independent reactions. *See* Dark reactions
Limit dextrin, 450
Lineweaver-Burk plot, 229, 229f
α-1,6 linkages, 631–632, 631f
Linking number, 834–836, 834f
Linoleoyl CoA, 653, 653f
Lipases, 645–646, 646f
Lipid bilayers. *See also* Membrane lipids
  as cooperative structures, 348
  defined, 342, 348
  diagram, 348f
  formation of, 348
  as permeability barrier, 349–350, 350f
  self-assembly process, 348
  as self-sealing, 348
  stabilization of, 348
Lipid droplet, 645
Lipid rafts, 358
Lipid vesicles, 348–349, 349f
Lipids
  digestion of, 645–646, 646f
  energy storage in, 645
  fatty acids in, 342–344, 343f, 343t
  hydrolysis of, 647–648
  membrane. *See* membrane lipids
  metabolism of. *See* fatty acid metabolism
  transport of, 646–647, 646f
Lipinski's rules, 1036–1037, 1037f
Lipoamide, in citric acid cycle, 500–501, 501f
Lipoic acid, 499, 499f
Lipolysis, 648, 648f
Lipopolysaccharide (LPS), 982–983
Lipoprotein particles, 782
Lipoproteins, 781, 782f
  chylomicrons, 782, 782t, 783
  components of, 782
  high-density (HDLs), 782, 782t, 783, 787–788
  intermediate-density (IDLs), 782, 782t, 783
  low-density (LDLs), 782, 782t, 783–787, 783f, 786f
  metabolism, 783f
  very-low-density (VLDLs), 782, 782t
Liposomes
  defined, 348
  formation of, 348–349, 349f
  therapeutic applications of, 349
Lipoxygenase, 669, 669f
Live attenuated vaccines, 1006
Liver
  alcoholic injury of, 820–821
  in amino acid degradation, 687, 692–693
  in cholesterol synthesis, 779
  cirrhosis of, 821
  drug metabolism in, 1039
  ethanol metabolism in, 819–822
  fasting and, 817
  fatty, 820
  gene expression in, 942, 942t
  gluconeogenesis in, 478, 483–485, 818
  glucose 6-phosphate (G-6P) in, 622–623, 637
  glucose levels and, 622–623
  glucose sensor in, 636
  glycerol in, 648
  glycogen in, 618, 618f, 627
  glycogen metabolism in, 635–637, 636f
  glycogen phosphorylase in, 623–624, 624f
  glycolysis in, 472–473, 472f, 473f, 483–485
  ketone bodies in, 659, 659f
  ketone body synthesis in, 819, 819f
  peroxisomes in, 657f
  in starvation, 818–819
  in starved-fed cycle, 816–817
  in triacylglycerol synthesis, 769
L-lectins, 335
Local hormones, 669f
Lock-and-key model, 224, 224f
Long interspersed elements, 154
Long QT syndrome (LQTS), 388–389
Loops, 44, 44f
Lou Gehrig's Disease, 136
Lovastatin, 788, 788f, 1044, **1044**
Low-density lipoprotein receptor
  absence of, 784–785
  destruction of, 785–786
  LDL release, 785–786, 786f
  mutations in, 785–786
Low-density lipoproteins (LDLs), 782, 782t
  in cholesterol metabolism, 784
  cycling of, 787
  defined, 783
  in familial hypercholesterolemia, 784–785
  metabolism of, 783f
  receptor-mediated endocytosis of, 784, 784f
  schematic model of, 783f
LuxR, 934
Lyases
  defined, 666b
  reaction type, 245t
Lymphotactin, 55, 55f
Lynch syndrome, 851, 851f
Lysine
  acetylated, 952–953
  structure of, 33, 33f
Lysogenic pathway, 146
Lysosomes
  defined, 332
  function of, 332
  in I-cell disease, 332
  with lipids, 774, 774f
Lysozyme
  antibodies against, 990, 990f
  backbone model, 62f
  ball-and-stick model, 61f
  ribbon diagram, 62f
  space-filling model, 61f
Lytic pathway, 146

MacKinnon, Roderick, 380, 384
Mad cow disease, 56
Magnesium ions
  in ATP hydrolysis, 276–277
  in Calvin cycle, 592, 592f, 598–599
  in DNA cleavage, 269–271

Major histocompatible complex (MHC), 984–985, 996
Major histocompatible complex proteins, 995–1004
  anchor residues, 997–998, 997f
  class I, 996–997, 996f, 998, 1002–1003, 1003f
  class II, 1000–1003, 1001f
  diversity of, 1002–1003
  peptide presentation by, 996–998, 996f
  in plasma membrane, 996
  structure of, 996–998
  T-cell receptors and, 998–1003, 998f
  in transplant rejection, 1002–1003
Malaria, 206, 206f
Malate, 517, 517f
  in Calvin cycle, 600f
  in citric acid cycle, 508
  in gluconeogenesis, 480
Malate dehydrogenase, 508, 667
Malate synthase, 517
Malate-aspartate shuttle, 550, 550f
MALDI-TOF mass spectroscopy, 86, 86f
4-maleylacetoacetate, 704, **704**
Malic enzyme, 667
Malonyl ACP, 661, 663
Malonyl CoA, 662, 671
Malonyl transacylase, 663, 663t
Malonyl/acetyl transacylase (MAT), 664, 665f
Maltase, 324, 450
Maltose, 324, 324f
Mammalian liver homogenate, 852
Manganese, 574, 574f
Manganese center, 574, 574f
Mannose, 317–318, 318f, 320f
Mannose 6-phosphate, 332, 332f
Mannose 6-phosphate receptor, 334
Maple-syrup urine disease, 705
Mass spectroscopy
  MALDI-TOF, 86, 86f
  peptide sequencing by, 87–88, 88f
  in protein and peptide identification, 85–92
  proteomic analysis by, 91–92, 92f
  tandem, 87–88, 88f
Mass-to-charge ratio, 85
Mast cells, 307, 307f
Matrix-assisted laser desorption/ionization (MALDI), 85, 332
Maximal rate, 228–229, 230
McArdle disease, 637t, 638
Mechanism-based (suicide) inhibition, 238, 238f
Mediator, 946, 946f
Megasynthases, 670
Melanin, synthesis of, 734, 734f
Melanocyte-stimulating hormone (MSH), 805–806, 806f
Melting temperature
  of DNA, 116
  of fatty acid chains, 357, 357t
  of phospholipid membrane, 357, 357f
Membrane anchors, 354, 354f
Membrane channels
  gap junction, 368
  ion, 368, 378–389
  membrane permeability to, 390–391, 391f
  water, 390–391, 391f
Membrane diffusion
  facilitated, 367, 369
  lateral, 356, 357, 357f
  lipid, 356
  protein, 357
  simple, 368
Membrane lipids. *See also* Lipid bilayers
  as amphipathic molecule, 347
  archaeal, 346–347, 346f
  carbohydrate moieties, 345–346
  cholesterol, 346, 776–788
  defined, 342
  diffusion coefficient, 356
  fatting acids in, 342–344, 343f, 343t
  glycolipids, 344, 345–346
  hydrophilicity of, 347, 347f
  hydrophobicity of, 347, 347f, 348
  lateral diffusion, 356, 356f, 357, 357f
  metabolism, 775–776
  movement in membranes, 357, 357f
  permeability barrier, 350
  phospholipids, 344–345
  phospholipids, synthesis of, 768–776
  rate of travel, 356
  representations of, 347f
  sphingolipids, synthesis of, 772–774
  synthesis of, 767–797, 775f
  triacylglycerols, synthesis of, 768–776
  as two-dimensional solutions, 342
Membrane potential, 369
Membrane proteins
  α helix of, 351–352, 355–356, 355f, 355t
  ATP hydrolysis, 370–376
  β strands of, 352, 352f
  carriers, 367
  channels. *See* membrane channels
  content variation, 350
  defined, 342
  diffusion, 357
  function of, 350
  gel patterns of, 350, 350f
  hydropathy plots for, 355, 356f
  hydrophobicity of, 354, 354f
  integral, 351, 351f
  linking to membrane surface, 353
  membrane interaction, 351–354
  peripheral, 351, 351f
  pumps. *See* pumps
  as two-dimensional solutions, 342
Membrane transport
  ABC transporters in, 367, 374–376, 374f, 375f
  active, 367, 369
  free energy in, 369, 369f
  passive, 369
  potassium ion channels and, 383, 383f
  primary active, 367
  secondary active, 367
  transporters. *See* transporters
Membranes, 341–363
  archaeal, 346–347, 346f
  as asymmetric, 342, 357–358, 359f
  common features of, 342
  defined, 341
  as electrically polarized, 342
  endoplasmic reticulum (ER), 668
  enzymatic activities, 358–359
  fluidity of, 357–359
  ion gradients across, 433–434, 433f
  lipid bilayer, 342, 348–350
  mitochondria, 524–525
  as noncovalent assemblies, 342
  overview of, 341–342
  permeability of, 341, 349–350, 350f, 367
  planar bilayer, 349, 349f
  polarity scale for transmembrane helices, 355, 355t
  processes of, 350–356
  protein interaction, 351–354
  as sheetlike structures, 342
  synthesis of, 359
Membrane-spanning α helices, 351–352, 355, 355f
Memory B cells, 1006
Memory T cells, 1006
Menten, Maud, 226
6-mercaptopurine, 1051, **1051**
β-mercaptoethanol, 49, 49f, 50f
Mercury poisoning, 515–516
Merrifield, R. Bruce, 94
Meselson, Matthew, 115, 116
Metabolic pathways. *See also specific pathways*
  amphibolic, 425
  biosynthetic, 425
  classes of, 424
  defined, 424
  degradative, 425
  formation of, 426
  illustrated, 425f
  motifs, 435–445
  regulation of, 424
  thermodynamically favorable, 425, 426
  thermodynamically unfavorable, 426
Metabolic syndrome, 809–810, 810f
Metabolism, 423–446
  activated carriers in, 435–440
  anabolic reactions in, 428–429
  carbon bonds cleavage, 442, 442f
  catabolic reactions in, 428
  catalytic activity control and, 443–444
  cholesterol synthesis, 784
  common motifs in, 423–424, 435–445
  defined, 423
  drug, 1037–1039
  enzyme control and, 443
  evolution of, 444–445
  first-pass, 1039
  free energy in, 426–427
  fructose, 466, 466f

Metabolism—(*continued*)
  glucose, 424f
  of glucose 6-phosphate, 607–610
  group-transfer reactions in, 441, 441f
  hormones and, 443–444
  hydrolytic reactions in, 441–442, 442f
  integration of, 801–822
  interconnected interactions in, 424–426, 424f
  iron, 954–956
  isomerization reactions in, 441, 441f
  ligation reactions in, 440–441, 440f
  lipid, 775–776
  lipoproteins, 783f
  overview of, 423–424
  oxidation-reduction reactions in, 440, 440f
  of pyruvate, 462–464
  regulation of, 442–444
  in starvation, 818–819, 819t
  substrate accessibility control and, 444
  substrate cycles in, 485, 485f
  thermodynamics of, 425–426
  types of reactions in, 440t
Metabolites, 2, 650
Metal ion catalysis
  in ATP hydrolysis, 276–277
  in carbon dioxide hydrolysis, 265–267
  defined, 252–253
  in DNA cleavage, 271–272, 272f
Metal ions, **829**
Metalloproteases, 261f, **261,** 262
Metals, cofactor, 217, 217t
Metamorphic proteins, 55–56
Metarhodopsin, 972
Methionine
  degradation, 701
  from homocysteine, 725
  metabolism, 701, 702f
  oxidation of, 303, 303f
  structure of, 31f, 32
  in translation, 906, 906f
Methionine sulfoxide, 303, 303f
Methionine synthase, 725
Methotrexate, 757
Methylases, 274
Methylation
  of adenine, 274–275, 275f
  in amino acid synthesis, 724–726
  DNA protection by, 274–275, 274f
  in gene expression, 949
  of phospholipids, 726
  reaction, 654
Methylcobalamin, 725
β-methylcrotonyl CoA, 702
β-methylglutaconyl CoA, 702
Methylmalonyl CoA, 655–656, 655f, **656**
N-methylmesoporphyrin, 241, 241f
Metmyoglobin, 194
Metoprolol, 1050
Mevalonate, 776–777, 777f
Micelles, 348, 348f
Michaelis, Leonor, 226
Michaelis constant (K_M)
  in Cheng-Prusoff equation, 1036
  defined, 227
  determination of, 228–229
  as enzyme characteristic, 229–230
  physiological consequences of variations, 228
  values of enzymes, 229t
Michaelis-Menten equation, 229, 233
Michaelis-Menten kinetics, 225–233, 227f, 235, 254, 287
Microbiomes, 19, 19f
Microbodies, 593
MicroRNAs, 879, 879f, 956–957, 956f
Microsomal ethanol-oxidizing system (MEOS), 820
Microtubules
  arrangement of, 1023, 1023f
  in cilia, 1023
  defined, 1022
  dynamic instability of, 1023
  in flagella, 1023
  function of, 1023
  kinesin movement along, 1022, 1024–1026, 1025f
  structure of, 1022–1023, 1022f
  tubulins in, 1023, 1023f
Mineralocorticoids, 789
Minoxidil, 1038
Mismatch DNA repair, 848, 848f
Mitchell, Peter, 542, 542b, 558, 578
Mitochondria
  in apoptosis, 557–558
  ATP synthase in, 550–552
  brown fat, 555
  in citric acid cycle, 496, 496f
  codons of, 126, 126t
  electron entry via shuttles, 549–550
  electron micrograph and diagram of, 525f
  endosymbiotic origin of, 525–526
  genome of, 525–526, 525f, 526f
  matrix, 524
  membrane, 524
  overlapping gene complements of, 526, 526f
  oxidative phosphorylation in, 524–526
  properties of, 524–526
  structure of, 524–525, 525f
Mitochondrial ATP-ADP translocase, 550–551, 551f, 557
Mitochondrial ATPase. *See* ATP synthase
Mitochondrial biogenesis, 813, 813f
Mitochondrial diseases, 557
Mitochondrial mutants, 557
Mitochondrial outer membrane permeabilization (MOMP), 558
Mitochondrial porin, 524
Mitochondrial transporters, 551–552, 551f, 552f
  defined, 552f
  illustrated, 552f
  for metabolites, 551–552
  structure of, 551, 551f
Mixed-function oxygenase, 703, 790–791
Molecular evolution, 185–187
Molecular heterogeneity, 242–243, 242f
Molecular imaging agents, 406
Molecular models
  backbone, 61–62, 62f
  ball-and-stick, 22–23, 23f, 61, 61f
  proteins, 61–62
  small molecules, 22–23
  space-filling, 22, 23f, 61, 61f
Molecular motors, 1011–1031
  ATP binding to, 1014–1016, 1015f, 1016f
  dynein, 1012, 1014, 1015f, 1022–1026
  flagella, 1012, 1023, 1026–1029
  kinesin, 1012, 1014, 1022–1026
  myosin, 1012, 1013–1014, 1013f, 1014f, 1018–1022
  operation of, 1011–1012
  overview of, 1011–1012
  P-loop NTPase superfamily, 1012–1016
  structures of, 1012–1016, 1013f, 1014f, 1015f
Molecular oxygen, toxic derivatives of, 538–540
Molecules
  amphipathic nature of, 348
  Fischer projections, 22
  homologous, 170
  permeability of, 350, 350f
  stereochemical renderings, 22
  three-dimensional structure of, 22–23
Moloney murine leukemia virus, 159
Monoclonal antibodies
  defined, 80
  generation for proteins, 80–82
  illustrated, 80f
  preparation of, 81f
  in signal transduction inhibition, 416–417
Monod, Jacques, 198, 928
Monomers, 107
Monooxygenase, 703, 790–791
Monosaccharides. *See also* Carbohydrates
  alcohols and, 322
  aldoses, 316
  amines and, 322
  boat form of, 320–321, 321f
  chair form of, 320–321, 321f
  defined, 316
  diastereoisomers, 317, 317f
  envelope form of, 321, 321f
  glycosidic bonds, 322, 322f
  isomers of, 317, 317f
  ketoses, 316
  linkage of, 323–325
  modified, 322–323, 322f
  reducing sugars, 321–322
  ring forms of, 319–321, 319f, 320f
  stereoisomers, 317, 317f
  structure of, 316–318
Morphine, 1037
MotA-MotB pairs, 1027, 1028f
Motifs
  defined, 48
  helix-turn-helix, 48

metabolism, 423–424
repeating, 91, 91f
mRNA
cDNA prepared from, 149–150
complementarity with DNA, 122, 122f
defined, 105, 120
5′ cap of, 878–879, 879f
FMN and, 867
modification of, 123, 123f
poly(A) tail of, 878–879
precursors, 876, 880–881, 882–883, 882t
in protein synthesis, 126
RNA editing and, 879–880, 879f
structure of, 912, 912f
transcription and, 883–884, 884f
transferrin-receptor, 955, 955f
Mucins
defined, 326, 329
overexpression of, 329
structure of, 329, 329f
Mucolipidosis, 332
Mucopolysaccharidoses, 328
Mucus, 329
Muller, Hermann, 847
Multidrug resistance, 374
Multidrug resistance protein, 374
Multifunctional protein kinases, 295
Multiple myeloma, 81
Multiple-substrate reactions
classes of, 232
defined, 231–232
double-displacement, 232, 233
sequential, 232
Muscle
ATP in contraction, 469–470
excess fatty acids in, 810
fatty acids in, 810, 815
fiber types, 625, 625t
glycogen in, 618
glycogen phosphorylase in, 625, 625f
glycolysis in, 469–471
insulin resistance in, 810–812
myofibrils in, 1020
PP1 regulation in, 634, 634f
sarcomere in, 1020, 1020f
Muscle contraction
actin in, 1019–1021
alanine in, 486, 486f
ATP formation in, 813–815
ATP hydrolysis in, 1019–1022, 1020f
fuel sources for, 813–815, 814t, 815f
gluconeogenesis in, 485–487, 486f
lactate in, 485–486
myosin in, 1019–1021
sliding-filament model of, 1020–1021, 1021f
Muscle-relaxation pathway, 1042–1043, 1043f
Mutagenesis
cassette, 151, 151f
oligonucleotide-directed, 150, 151f
by PCR, 151–152, 151f
site-directed, 150–151, 151f, 260, 260f
Mutagens, 846–847
Mutase, 460
Mutations
deletion, 150, 151–152, 151f
disease-causing, functional effects of, 152
in genes encoding hemoglobin subunits, 204
insertion, 150, 151
at level of function, 177–178
at level of sequence, 177
point, 150
in recombinant DNA technology, 150–152
substitution, 150
MWC model, 198–199, 199f
Mycobacterium tuberculosis, 142, 687
Myofibrils, 1020
Myogenin, 160, 161f
Myoglobin
alignment comparison, 173–174, 173f
alignment with gap insertion, 173, 173f
amino acid sequences of, 171–172, 171f, 172f
as compact molecule, 46
defined, 46, 171, 191
distribution of amino acids in, 47f
hemoglobin comparison, 191–192
Hill plot, 210f
oxygen-binding, 192–195, 196f
oxygen-binding curve, 196, 196f
as single polypeptide, 192
structure of, 192, **192,** 194–195
tertiary structure of, 178–179
three-dimensional structure of, 46f
Myosin
active site, 279, 279f
altered conformation persistence, 278–279
ATP binding to, 1014–1016, 1015f, 1016f
ATP complex structure, 276, 276f
ATP hydrolysis and, 276–278
ATPase transition-state analog, 277, 277f
binding to actin filaments, 1018, 1018f
catalytic strategies of, 251–252
conformational changes, 278, 278f
defined, 275, 1012
heavy chains, 1012, 1014, 1014f
kinetic studies of, 276
lever arm, 1015, 1015f, 1022, 1022f
light chains, 1013, 1013f
motion observation, 1018–1019, 1019f
movement along actin, 1016–1022, 1020f
in muscle contraction, 1019–1021
P-loop NTPase, 1012
P-loop structures, 280–281
power stroke release, 1019
regulatory light chain, 1012
relay helix, 1016, 1016f
S1, 1013, 1013f, 1015, 1015f
S2, 1013–1014, 1013f
single-molecule motion, 279–280, 280f, 1018–1019, 1019f
as slow enzymes, 278
structure of, 1013–1014, 1013f
switch I and switch II, 1015–1016, 1016f
in thick filaments, 1020, 1021f
in thin filaments, 1020
Na^+ ions. *See* Sodium ions
Na^+-K^+ pumps, 370, 373
NAD^+ (nicotinamide adenine dinucleotide)
binding site for, 465, 465f
in glycolysis, 458, 459f, 462–464, 465
light absorption and, 67
from metabolism of pyruvate, 462–464
reduction potential of, 528
regeneration of, 464, 464f
structures of oxidized form of, 435, 435f
synthesis of, 734, 734f
NADH (nicotinamide adenine dinucleotide reduced)
in ATP generation, 436
electrons from, 549–550
electron-transport chain and, 497, 510, 526, 528–529
ethanol metabolism in, 820–821
in fatty acid oxidation, 650–652
in glycolysis, 458, 462–464
light absorption, 67
pyruvate reduction by, 464
in respiratory chain, 532–533, 533f
structures of oxidized form of, 435, 435f
transport of, 549–550
NADH-cytochrome b_5 reductase, 668, 668f
NADH-Q oxidoreductase, 529–531, 530f, 530t, 532–533, 533f
$NADP^+$ (nicotinamide adenine dinucleotide phosphate), 602, 607
NADPH (nicotinamide adenine dinucleotide phosphate reduced)
in Calvin cycle, 589–590, 597
dihydrofolate reductase and, 756
in fatty acid synthesis, 667
formation by photosystem I, 579
in glucose 6-phosphate conversion, 602
pathways requiring, 601f
in pentose phosphate pathway, 608–610
in reductive biosynthesis, 436
Nalidixic acid, 839
Neanderthals, 185, 185f
Neck linker, 1016
Negatively charged amino acids, 33–34, 34f
Neher, Erwin, 379
Nernst equation, 387
Nerve impulses, 378, 385
Neural-tube defects, 762
Neuroglobin, 208
Neurological diseases. *See* Diseases and disorders
Neuropeptide Y (NPY), 805–806, 806f
Neurotransmitter release, 360, 361f
Neurotransmitters, 385
Next-generation sequencing (NGS)
defined, 155
detection methods in, 155, 155f
as highly parallel, 155
ion semiconductor, 156
pyrosequencing, 155–156
reversible terminator method, 155
NF-κB, 687
9 + 2 array, 1023

Nirenberg, Marshall, 124
Nitric oxide (NO), synthesis of, 735, 735f
Nitric oxide synthase, 735
Nitrogen
  in amino acids, 714
  excess, disposal of, 693–698
  metabolism, 696f, 733
  peripheral tissue transport of, 692–693
Nitrogen fixation
  in amino acid synthesis, 714–718
  ATP in, 715–718
  defined, 714
  FeMo cofactor in, 716, 716f
  glutamate in, 717–718
  glutamine in, 717–718
  iron-sulfur proteins in, 715–716, 715f
  P clusters in, 716, 716f
Nitrogenase, 715, 716, 716f, **716**
Nitrogenase complex, 715
^{15}N labeling, 735b
N-linkage, 326, 326f
N-linked oligosaccharides, 326, 326f, 332
Nociceptors, 977
NompC, 976
Noncompetitive inhibition. *See also* Enzyme inhibition
  defined, 234
  double-reciprocal plot, 236, 236f
  example of, 236
  inhibitor binding, 236
  kinetics of, 235–236, 235f
Nonconservative substitutions, 174, 175f
Noncovalent bonds
  in biological molecules, 7–8
  hydrogen bonds, 5, 7, 8, 8f, 9, 10
  hydrophobic effect and, 9–10, 9f
  types of, 7
  van der Waals interactions, 7, 8, 8f, 10
Nonessential amino acids
  defined, 719
  synthesis of, 719, 720
Nonheme iron proteins, 531
Nonhomologous end joining (NHEJ), 849
Nonribosomal peptides, 670
Nonshivering thermogenesis, 555f, 556, 558
Northern blotting, 138
Novobiocin, 839
N-terminal rule, **684**
NTPases, P-loop. *See* P-loop NTPases
Nuclear envelope, 360, 360f
Nuclear hormone receptors
  defined, 949
  domain structures, 950, 950f
  drug-binding to, 951–952
  in gene expression, 949–951
  ligand-binding to, 950, 951f
Nuclear magnetic resonance (NMR) spectroscopy, 65, 97–100
  basis of, 98, 98f
  chemical shifts, 98
  defined, 97
  NOESY and, 99–100, 99f, 100f
  one-dimensional, 98, 98f
  RNA structure, 187, 187f
  signals, 98f
Nuclear Overhauser effect (NOE), 99
Nuclear Overhauser enhancement spectroscopy (NOESY), 99–100, 99f, 100f
Nuclear pores, 360
Nucleation-condensation model, 53
Nucleic acids. *See also* DNA; RNA
  backbones in, 106–107, 106f
  bases, 106–109
  complementary sequences, 109–114
  complex structures, 113–114, 113f, 114f
  defined, 105
  melting of, 116
  nucleotides as monomeric units of, 107–108
  polymeric structure, 106f
  solid-phase synthesis of, 136
Nucleoside 5′-phosphate, 108
Nucleoside diphosphate, 245
Nucleoside diphosphate kinase, 427, 747
Nucleoside diphosphokinase, 506
Nucleoside monophosphate, 427
Nucleoside monophosphate kinases, 245, 427, 747
Nucleoside phosphorylases, 760, 761
Nucleosides
  defined, 107, 744
  nomenclature of, 744
  units, 107–108
Nucleosome core particle, 943, 943f
Nucleosomes, 943–944
Nucleotidases, 760
Nucleotide synthesis, 743–763
  de novo pathways in, 744–752, 744f
  deoxyribonucleotide, 753–758, 759–760, 759f
  disorders of, 760–762
  inosinate in, 759
  overview of, 743–744
  of purine, 748–752, 748f, 758–759, 758f
  of pyrimidine, 744–748, 744f, 758
  regulation of, 758–760
  salvage pathways in, 744–752, 744f
  substrate channeling in, 745–746, 746f
Nucleotide triphosphates, 107
Nucleotides
  cyclic, 743
  defined, 106, 743, 744
  DNA, 107
  in encoding amino acids, 124
  ester linkage, 107
  metabolism disruptions, 760–762
  as monomeric units of nucleic acids, 107–108
  nomenclature of, 744
  RNA, 107–108
  sequences in DNA, 135–136
  in signal transduction pathways, 743
Nutrition, 20, 20f

Obesity, 801–812
  causes of, 803–804
  diabetes and, 807–812
  dieting and, 807
  evolution and, 802–804
  health consequences of, 803t
  insulin in, 805–806
  insulin resistance and, 809–810, 810f
  leptin resistance in, 806–807, 806f
  overview of, 801–802
  prevalence of, 803–804
Obligate anaerobes, 464, 465t
Octadecanoic acid, 342
Odd-chain fatty acids, 654, 654f
Odorant receptors (ORs), 964f
  activation of, 965f
  conserved and variant regions, 963–964, 964f
  evolution of, 963–964, 963f
  in olfaction, 963–965
Off-diagonal peaks, 99
Okazaki, Reiji, 831
Okazaki fragments, 831–832, 831f, 841
Olfaction, 962–966. *See also* Sensory systems
  anosmias, 962
  combinatorial mechanisms in, 964–966
  electronic nose in, 965–966, 966f
  evolution of, 963–964, 963f
  impaired, 962
  main olfactory epithelium in, 962, 963f
  molecule shape and, 962
  odorant receptors (ORs) in, 963–965, 963f, 964f
  7TM receptors in, 963
  signal transduction in, 964, 964f
Olfactory neurons, 964, 965f
Oligomeric 7TM receptors, 969
Oligomerization, in B-cell activation, 993–994
Oligomycin, 556
Oligonucleotide-directed mutagenesis, 150, 151f
Oligopeptides. *See* Peptides
Oligosaccharide chain, 772
Oligosaccharides, 332f
  antigen structures, 331, 331f
  attached to erythropoietin, 327, 327f
  defined, 323
  enzymes responsible for assembly, 331
  N-linked, 326, 326f, 332
  O-linked, 326
  sequencing of, 332–333, 333f
O-linkage, 326
Omeprazole (Prilosec), 1049, **1049**
Oncogenes, 416
One-dimensional NMR, 98, 98f
Open promoter complex, 866, 866f
Open reading frames (ORFs), 177
Operator site, 928
Operons, 929f, 930f
  defined, 928–929
  histidine, 936–937, 936f

  lac, 929–932, 929f, 930f
  structure of, 928, 929f
  threonine, 936, 936f
  trp, 936
Opsin, 970, 971
Optical traps, 1019f
Oral bioavailability, 1036, 1037f
Origin of replication complexes, 843
Origin of replication (oriC locus), 842
Ornithine, 694–695
Ornithine transcarbamoylase, 694, 696
Ornithine transcarbamoylase deficiency, 697–698, 698f
Orotate, 746–747, 746f, 747f
Orotate phosphoribosyltransferase, 747
Orthologs, 170, 171f
Orthophosphate, 619
Osteoarthritis, 328
Osteogenesis imperfecta, 46
Osteomalacia, 795
Otto, John, 307b
Overlap peptides, 89, 89f
Oxaloacetate, **233**
  in amino acid degradation, 698, 700
  in Calvin cycle, 600, 600f
  in citric acid cycle, 496, 502, 507–508
  conversion into phosphoenolpyruvate, 480
  in fasting or diabetes, 657
  formation of, 480
  in gluconeogenesis, 477f, 478–480, 480f, 819
  in glyoxylate cycle, 517f
  synthesis of, 440–441, 441f
Oxidation
  activated carriers of electrons for, 435
  of aldehydes, 457
  carbon, 432–433
  in drug metabolism, 1037, 1039
  energy, 433
  in fatty acid metabolism, 648–649, 650–652, 651t
  of glyceraldehyde 3-phosphate (GAP), 458–459, 459f
  of membrane proteins, 652f
  phase I transformations in, 1039
  pyruvate dehydrogenase complex and, 498–499
  of succinate, 507–508
Oxidation-reduction potential. *See* Reduction potential
Oxidation-reduction reactions
  free-energy change in, 528
  in metabolism, 440, 440f
Oxidative phosphorylation, 523–560
  ADP in, 550–551, 551f, 553–554, 554f
  ATP yield in, 552–553, 553t
  ATP-ADP translocase in, 550–551, 551f, 557
  cellular respiration in, 552–558
  chemiosmotic hypothesis in, 542–543, 542f, 543f
  defined, 433, 523
  electron transfer in, 526–529
  electron-transport chain in, 548f
  in exercise, 814–815
  inhibition of, 556
  iron sulfur clusters in, 531–532, 532f
  in mitochondria, 524–526, 525f, 526f
  mitochondrial transporters in, 551–552, 551f, 552f
  NADH in, 526, 528–529
  overview of, 523–524, 524f, 548f
  photosynthesis comparison, 579f
  proton gradients in, 433, 433f, 524, 529, 541–549
  proton-motive force in, 523, 542–543, 542f, 544–545, 545f
  rate of, 553–554, 554f
  reduction potential in, 526–528, 526f
  respiratory chain in, 433, 529–541
  respiratory control in, 553–554, 554f
  shuttles in, 549–552, 549f, 550f
  transporters in, 549–552, 551f, 552f
  uncoupling of, 554–556, 555f
Oxidized lipoamide, 499–500
Oxidoreductases, 245t
Oxidosqualene cyclase, 778
Oxoguanine-adenine base pair, 846f
Oxyanion hole
  of chymotrypsin, 257, 257f
  defined, 257
  of subtilisin, 259, 259f
Oxygen
  affinity in red blood cells, 200–201, 201f
  bound, stabilizing, 194f
  carbon monoxide and, 201–202
  concentration in tissues, 197, 197f
  cooperative release of, 197
  evolution of, 574b
  in hydroxylation function, 790–791
  partial pressure, 194–195
  photons for generation of, 574, 574f
  reactive, 194
  toxic derivatives of, 538–540
Oxygen binding
  with concerted model, 212f
  cooperative, 196–197
  heme in, 193–194
  hemoglobin, 192–195, 196f
  hemoglobin quaternary structure and, 197–198, 198f
  myoglobin, 192–195, 196f
  pure hemoglobin, 200, 200f
  sites, 197
Oxygen therapy, hyperbaric, 202
Oxygenase reaction, 593–594, 593f
Oxygenases
  in aromatic amino acid degradation, 703–705
  mixed-function, 703
Oxygen-binding curve
  defined, 195
  fractional saturation, 195
  for hemoglobin, 196, 196f
  for Hill coefficients, 210–211, 211f
  for myoglobin, 196, 196f
  sigmoid, 196
  T-to-R transition, 199, 199f
Oxyhemoglobin, 198
Oxymyoglobin, 192, **193**

P clusters, 716, 716f
P site, ribosomal, 905, 905f
p53, 851
p160 family, 951
P680, 572, 573
P700, 572, 576, 576f
P960, 571–572
PALA, 289–290, 289f, **289,** 290f
Palindromes, 137, 137b
Palmitate, 652, 652f, 666
Palmitoleate, 652–653, 652f
Palmitoyl CoA, 652, 652f, 671, 772
Pamaquine, 611
Pancreas
  α cells, 802
  β cells, 804
  gene expression in, 942, 942t
Pancreatic failure, 810–812
Pancreatic lipases, 645–646, 646f
Pancreatic trypsin inhibitor, 302–303, 302f
Pancreatic zymogens, 301–303, 301f
Papain, 216–217
Paracoccus denitrificans, 541
Paralogs, 170, 171f
Paraquat, 584
Park, James, 1042
Parkinson disease, 56–57, 683, 685
Partial pressure, oxygen, 194–195
Passenger strand, 162
Passive membrane transport, 369
Pasteur, Louis, 450, 1006
Patch-clamp technique, 379, 379f
Pathogen-associated molecular patterns (PAMPs), 982–983, 983t, 984f
Pauling, Linus, 193, 204–205, 222b, 240
PCSK9 (proprotein convertase subtilisin/kexin type), 787, 788
Penicillin, 1041–1042, 1041f, **1041**
  conformations of, 240, 240f, **240**
  mechanism of action, 239–240, 240f
  reactive site of, 239, 239f
  thiazolidine ring, 239
Penicillium citrinum, 1044
Penicilloyl-enzyme derivative, 240, 240f
Pentose phosphate pathway, 601–612
  Calvin cycle and, 610
  in cell growth, 610
  defined, 589, 601
  erythrose 4-phosphate in, 604, 604f
  fructose 6-phosphate in, 604, 604f
  glucose 6-phosphate in, 602, 607–610
  glucose 6-phosphate dehydrogenase in, 610–612
  glyceraldehyde 3-phosphate (GAP) in, 603, 604, 604f
  glycolysis and, 602–610
  illustrated, 602f
  modes, 608–609

Pentose phosphate pathway—(*continued*)
  NADP$^+$ in, 607
  NADPH in, 608–610
  nonoxidative phase of, 602–607, 605t
  overview of, 589–590
  oxidative phase of, 602, 603f, 605t
  rate of, 607
  reactions, 605–607, 605f, 606f
  ribulose 5-phosphate in, 602, 603, 603f, 604, 608–609
  sedoheptulose 7-phosphate in, 603, 603f, 604f
  tissues with, 609t
  transaldolase in, 602–605, 606–607, 606f
  transketolase in, 602–606, 605f
Pentose shunt. *See* Pentose phosphate pathway
Pentoses, 316
Peptide bonds
  chemical nature of, 253
  cis configuration, 38, 38t, 39t
  cleavage of, 59, 258
  defined, 35
  formation in translation, 908–909, 908f
  formation of, 35f
  hydrolysis of, 253
  lengths, 38, 38t
  as planar, 38, 38t
  rigidity of, 39
  rotation of, 38–39, 39t
  stability of, 35
  torsion angles, 39, 39b
  trans configuration, 38, 38t, 39t
  X-Pro linkages, 38, 39t
Peptide mass fingerprinting, 91
Peptide sequences, 52f
Peptide-ligation methods, 95
Peptides
  gastrointestinal, 804–805, 805t
  mass spectroscopy for sequencing, 87–88, 88f
  nonribosomal, 670
  overlap, 89, 89f
  presentation by internalized proteins, 1000
  presentation by MHC proteins, 996–998, 996f
  synthesis of, 92–95, 94f
Peptidoglycan, 239, 239f
Peptidyl transferase center, 908
Perforin, 1000
Peripheral membrane proteins, 351, 351f
Periplasm, 359
Periplasmic side, 570
Permeability
  lipid bilayers, 349–350, 350f
  membrane, 341, 367, 390–391
  of potassium ion channels, 382
  selective, 341
Permeability coefficients, 350, 350f
Perovskite, 585
Peroxide bridge, in cytochrome *c* oxidase, 538, 538f
Peroxisomes
  in Calvin cycle, 593
  defined, 657
  electron micrograph of, 594f, 657f
  in fatty acid oxidation, 656–657, 657f
Perutz, Max, 195
PEST sequences, 684
P-glycoprotein, 374
pH
  buffers and, 15–17
  carbon dioxide and, 203, 204
  defined, 13
  effect on carbonic anhydrase activity, 265, 265f
  ionization state as function of, 30f
Pharmacogenetics/pharmacogenomics, in drug development, 1050
Pharmacology, 1034b. *See also* Drug development
Phase I transformations, in oxidation, 1039
Phase II transformations, in conjugation, 1039
Phenotype-genotype correlation, 1050–1051, 1050f
Phenyl isothiocyanate, 87
Phenylalanine
  degradation of, 703, **703,** 704f
  structure of, 31f, 32
  synthesis of, 727–728, 728f
Phenylalanine hydroxylase, 703
Phenylketonuria, 705
Phenylpyruvate, 728
Phosphatases
  deficiency, 512
  defined, 295
  in phosphorylation, 294–296, 295t
  PP2A, 296
Phosphate carrier, 552
Phosphate esters, 434
Phosphates, 16, 434
Phosphatidate, **344,** 769, **769**
Phosphatidic acid phosphatase (PAP), 768, 775–776, 775f
Phosphatidylcholine, synthesis of, 770–771, 771f
Phosphatidylethanolamine, 770, **770**
Phosphatidylethanolamine methyltransferase, 771
Phosphatidylinositol, 769, **769**
Phosphatidylinositol 3,4,5-trisphosphate (PIP$_3$), 410, 410f, 808–809
Phosphatidylinositol 4,5-bisphosphate (PIP$_2$), 404, 405f, 410, 410f, 770, 808
Phosphatidylinositol-dependent protein kinase (PDK), 809
Phosphatidylserine, 771
Phosphocreatine, **232**
Phosphoenolpyruvate (PEP)
  formation of, 461
  in gluconeogenesis, 477f
  in glycolysis, 461
  NAD$^+$ regeneration in, 462–464
  phosphoryl-transfer potential, 430, 461
Phosphoenolpyruvate carboxykinase, 480
Phosphoenolpyruvate carboxylase, 600
Phosphofructokinase (PFK), 454
  activation of, 472, 472f
  in liver, 472, 472f
  in muscle, 469–470, 470f
  regulation of, 472, 472f
Phosphofructokinase 2 (PFK2), 483–484, 484f
Phosphoglucomutase, 623f
Phosphogluconate pathway. *See* Pentose phosphate pathway
6-phosphogluconate dehydrogenase, 602
6-phosphoglucono-δ-lactone, 602, 603f
Phosphoglucose isomerase, 454
2-phosphoglycerate, in glycolysis, 460–461, 460f
3-phosphoglycerate
  in Calvin cycle, 591–592, 592f
  in cysteine synthesis, 722–723
  in glycine synthesis, 722–723, 723f
  in glycolysis, 460–461, 460f
  in oxygenase reaction, 593, 593f
  in serine synthesis, 722–723, 722f
3-phosphoglycerate dehydrogenase, 730–731
Phosphoglycerate kinase, 459
Phosphoglycerate mutase, 460, 622
Phosphoglycerides, 344
Phosphoglycolate, 593, 593f, 594f
Phosphoinositide cascade, 629
  activation of, 404–405
  defined, 404
  illustrated, 405f
  reaction, 405, 405f
Phospholipase C, 404
Phospholipid synthesis
  from activated alcohol, 770
  activated intermediate, 769–770
  phosphatidylcholine, 770–771, 771f
  sources of intermediates in, 768–776, 768f
Phospholipids
  base-exchange reactions in generation of, 771
  in bimolecular sheet formation, 348–350
  flip-flop, 357, 357f
  lipid vesicles formed from, 348–349
  melting temperature, 357, 357f
  membrane, 344–345
  platform for, 344
Phosphopantetheine, 662f
Phosphopentose epimerase, 595
Phosphopentose isomerase, 595
Phosphoribomutase, 760
5-phosphoribosyl-1-pyrophosphate (PRPP)
  in purine synthesis, 749, 759
  in pyrimidine synthesis, 728, 746–747, 747f
Phosphorolysis, in glycogen metabolism, 619–623
Phosphorothioates, 270–271, 271f

Phosphoryl group, 296
Phosphorylase. *See* Glycogen phosphorylase
Phosphorylase *a*, 623–624, 624f, 626, 627f, 636, 636f
Phosphorylase *b*, 623, 624f, 625, 626, 627f
Phosphorylase kinase
  activation of, 626, 627f
  defined, 626
  subunits, 626, 627f
Phosphorylation
  amplified effects of, 296
  ATP formation by, 580, 580f, 581
  cAMP stimulation of, 403
  of carboxyl-terminal domain (CTD), 875
  of carboxyl-terminal tail, 413
  in citric acid cycle, 511–512
  control of, 294–296
  in covalent modification, 293, 294t
  dephosphorylation and, 296
  EGF receptor, 413
  in enzyme regulation, 293–298, 294t
  free energy of, 296
  of fructose 6-phosphate, 454
  in gluconeogenesis, 480, 484
  in glycogen metabolism, 619–620
  in glycolysis, 451–453, 454
  in metabolism, 427, 433
  of nucleoside monophosphates, 427
  oxidative. *See* oxidative phosphorylation
  in phosphorylase activation, 626, 627f
  in phosphorylase conversion, 626
  in purine nucleotide synthesis, 749–751, 749t, 750f
  of pyruvate dehydrogenase, 511
  reactive intermediates, 323
  of serine residue, 484
  substrate-level, 460
  sugars, 323
Phosphorylethanolamine, 770, **770**
Phosphoryl-transfer potential, 429, 430–431, 430f, 432–433
Phosphoserine, **57,** 58
Phosphothreonine, 58
Phosphotriesters, 139, 140f, **140**
Photoinduced charge separation, 569–572, 569f
Photophosphorylation, cyclic, 580, 580f, 581
Photoreceptors, 970, 973–974, 973f
Photorespiration
  in C_4 pathway, 599–601
  in Calvin cycle, 594, 594f
Photosynthesis, 565–586
  accessory pigments in, 581–584
  artificial photosynthetic systems and, 585
  in autotrophs, 566
  bacterial reaction center, 569–572, 570f, 571f
  basic equation of, 565
  chlorophylls in, 568–572
  in chloroplasts, 567–568, 567f
  components of, location of, 583–584, 584f
  cyclic photophoshorylation in, 580, 580f, 581
  dark reactions, 566, 589
  defined, 565
  electron transfer in, 568–572
  energy conversion in, 566–567
  enzymes in, 217
  herbicide inhibition of, 584, **584**
  in heterotrophs, 566
  light absorption in, 569, 569f
  light energy conversion in, 566–567
  light reactions in, 566, 566f, 581, 584, 589
  light-harvesting complexes in, 583, 583f
  overview of, 565–567
  oxidative phosphorylation comparison, 579f
  photo charge separation in, 569–572, 569f
  photosystem I. *See* photosystem I
  photosystem II. *See* photosystem II
  proton-motive force in, 578
  reaction centers in, 569–572, 570f, 571, 572f, 576, 576f, 581–584
  resonance energy transfer in, 582–583, 582f
  in thylakoid membranes, 567–568
  yield in, 566b
  Z scheme of photosynthesis, 577, 577f
Photosynthetic catastrophe, 567b
Photosynthetic systems, 584–585, 584t
Photosystem I. *See also* Photosynthesis
  defined, 567, 572
  electron flow through, 580
  ferredoxin generated by, 576–577
  illustrated, 572f
  inhibitors of, 584
  link to photosystem II, 575, 575f
  NADPH formation by, 579
  reaction center, 572, 576, 576f
  structure of, 575–576, 575f
  Z scheme of photosynthesis, 577, 577f
Photosystem II. *See also* Photosynthesis
  defined, 567
  electron flow through, 573b
  electron source, 573–574
  illustrated, 572f
  inhibitors of, 584
  link to photosystem I, 575, 575f
  location of, 574–575
  photochemistry of, 573
  reaction center, 572, 573
  structure of, 573, 573f
  Z scheme of photosynthesis, 577, 577f
Phytol, 569
Pichia pastoris, 78
Ping-pong reactions, 233
Pinopsin, 974
pK_a values, 34t
Placebo effect, 1052
Planar bilayer membrane, 349, 349f
Planck's constant, 222
Plants
  C_3, 601
  C_4, 600–601
  genetically engineered, 163–164, 163f
  glyoxylate cycle in, 516–517, 517f
  oxaloacetate synthesis, 660
  photosynthesis in. *See* photosynthesis
  starch in, 597
  sucrose synthesis in, 597–598, 598f
  thylakoid membranes of, 583
Plaques, in genomic library screening, 148, 148f
Plasma-membrane proteins, 916, 916f
Plasmids
  defined, 144
  engineered, 145
  expression vectors, 146
  genes, 144
  microinjection of, 159f, 160
  polylinker region, 145, 145f
  Ti, 163–164, 163f
  as vectors for DNA cloning, 144–146
Plasmin, 308
Plasminogen, 308
Plasmodium falciparum, 206, 336, 525
Plastocyanin, 583
Plastoquinone, 572–573, 583
Pleckstrin homology domain, 410
P-loop NTPases
  in helicases, 832
  molecular motors, 1012–1016
  proteins containing, 280–281
P-loop structures, 280–281
PLP-Schiff-base linkage, 621
Pluripotent stem cells, 947
Point mutation, 150
Poisoning
  arsenite, 516, 516f
  mercury, 515–516
  ricin, 915
Polar amino acids
  structure of, 33
  types of, 32–33
Poly(A) tail-binding protein I (PABPI), 912, 912f
Polyacrylamide gel electrophoresis (PAGE), 72, 72f
Polyampholytes, 73
Polyclonal antibodies, 80, 80f
Polygalacturonase, 164
Polyketides, 670
Polylinker region, plasmid, 145, 145f
Polymerase chain reaction (PCR)
  cycle steps, 141f, 142, 142f
  defined, 141
  DNA sequence amplification by, 141–142, 141f, 142f
  first cycle of, 141f
  as gene exploration tool, 136
  inverse, 151, 151f
  multiple cycles of, 142f
  mutagenesis by, 151–152, 151f
  quantitative (qPCR), 157–158, 158f
  uses of, 142–143
Polymerase switching, 843–844
Polymerization reactions, 1018
Polymorphisms, 1002–1003, 1003f

Polypeptide chains
  backbone, 36, 36f
  bond rotation in, 38–39, 39t
  cleavage of, 88, 89t
  components of, 36f
  cross-linked, 36–37, 37f
  directionality of, 35–36
  disulfide bonds of, 36–37, 37f
  flexibility of, 38–39
  formation of, 35
  loops, 44, 44f
  random-coil conformation, 50
  residue, 35
  reverse turns, 44, 44f
  side chains, 36, 36f, 47
  subunit structures, 48–49
Polypyrimidine tract, 880
Polysaccharides
  defined, 324
  glycosidic bonds, 325, 325f
Polysomes, 910, 910f
Polyunsaturated fatty acids, 661, 669
Pompe disease, 637–638, 637t, 638f
Porin
  amino acid distribution in, 47f
  amino acid sequence of, 352, 352f
  defined, 352
  hydropathy plot for, 355, 356f
  structure of, **352**
Porphobilinogen, 736, 737, 737f
Porphobilinogen deaminase, 736
Porphyrias, 737–738
Porphyrins, 241
  disorders of, 737–738
  synthesis of, 736–737, 737f
Porter, Rodney, 985
Positively charged amino acids, 33, 33f
Posttranscriptional gene expression
  in eukaryotes, 954–957
  in prokaryotes, 935–937
Posttranslational modifications, 90
Potassium ion channels
  as archetypical structure, 379–380
  hERG, 389
  inactivation of, 384–385, 385f
  path through, 381, 381f
  permeability of, 382
  purification of, 380
  selectivity filter, 381, 381f, 382, 382f
  structure of, 380–382, 381f, 382
  transport model, 383, 383f
  voltage-gated, 383–384, 384f
Potassium ions
  action potentials and, 378
  dehydration of, 382, 382f
  hydration of, 383f
Potential energy, 11
Power stroke, 1019
PP2A, 296
Precursor ions, 87
Pregnenolone
  androgen synthesis by, 792–793
  corticosteroids from, 792
  defined, 792
  estrogen synthesis by, 792–793
  progesterone from, 792
  structure of, **793**
  synthesis of, 792–793
Pre-mRNA processing, 877, 877t, 878–879, 878f, 884–885, 884f
Prephenate branch, 728
Prilosec (omeprazole), 1049, **1049**
Primaquine, 611
Primary active transport, 367
Primary antibody, 83–84
Primary messenger, 398
Primary protein structure, 27, 35–40, 48, 78–79
Primase, 831
Primers
  defined, 829b
  DNA probes as, 140
  in DNA replication, 831, 831f
  match stringency, 142
  in polymerase chain reaction (PCR), 141–142, 141f
  RNA, 841
Prion diseases, 56–57, 57f
Prions, 56, 56f
Probes. *See* DNA probes
Procaspases, 299
Processivity
  defined, 621, 839, 840b
  in DNA replication, 839–840, 839f
  kinesin motion, 1024–1026
Procollagen, 300
Product ions, 87
Progesterone, 789, 792, **793**
Progestogen, 789
Programmed cell death. *See* Apoptosis
Proinsulin, synthesis of, 150, 150f
Prokaryotes. *See* Bacteria
Prokaryotic RNA polymerases, 860, **860**
Proliferating cell nuclear antigen (PCNA), 832
Proline
  degradation, 701, 701f
  structure of, 31f, 32
  synthesis of, 722–723
Prolyl hydroxylase, 821
Prolyl hydroxylase 2, 513
Promoter sites
  base sequences, 122b
  defined, 122
  for transcription, 122, 122f
Promoters
  alternative sequences, 865, 865f
  bacterial, 864, 864f
  closed complex, 865–866, 866f
  core, 864
  defined, 864
  open complex, 866, 866f
  strong, 864
  upstream of, 864
  weak, 864
Pro-N-terminal degrons, 684
Proofreading
  in DNA repair, 847, 848f
  in transcription, 863
  in translation, 900–901
Propionyl CoA, 654, 654f, 701–702
6-*n*-propyl-2-thiouracil, 967, **967**
Prostacyclin, 669
Prostacyclin synthase, 669, 669f
Prostaglandin, 669
Prostaglandin H_2
  attachment of, **353**
  defined, 353
  formation of, 353, 353f
  synthase-1, 353f, 354f
Prostaglandin H_2 synthase-1, 353–354, 353f
Prosthetic groups, 46, 217
Protease inhibitors
  α_1-antiproteinase, 303
  drugs, 263–264, 263f
  in HIV infection treatment, 263, 263f, 792, 1046, 1046f, 1053
  initial design of, 1046, 1046f
  pancreatic trypsin inhibitor, 302–303, 302f
  serine, 302
Proteases, 333
  active sites, 261f
  applied to glycoproteins, 333
  aspartyl, 261f, **261,** 262, 262f
  catalytic triad in, 258–260
  cysteine, 261–262, 261f, **261**
  metalloproteases, 261f, **261,** 262
  reaction facilitation, 253–264
  serine, 251
Protease-substrate interactions, 258, 258f
Proteasomes. *See also* Amino acid degradation
  19S regulatory unit, 685, 685f
  20S, 685, 685f
  26S, 685, 685f
  in amino acid degradation, 685–686
  defined, 681
  evolution of, 686, 686f
  free amino acid generation, 686f
  prokaryotic, 686, 686f
Protein colipase, 646
Protein Data Bank, 100
Protein domains, 48, 48f
Protein folding, 904f
  essence of, 53
  funnel, 54, 54f
  as highly cooperative process, 52–53
  illustrated, 21f
  Levinthal's paradox in, 53
  misfolding and, 811
  nucleation-condensation model, 53
  pathway of chymotrypsin inhibitor, 54f
  by progressive stabilization, 53–54
  transition from folded to unfolded, 52f
  typing-monkey analogy, 53, 53f
Protein identification
  cleavage in, 88–89, 89f
  genomic and proteomic methods as complementary, 89–90

  - MALDI-TOF mass spectroscopy in, 86, 86f
  - mass spectroscopy in, 85–92, 86f, 88f, 92f
  - peptide mass fingerprinting, 91
  - in protein studies, 65
- Protein kinase A (PKA)
  - bound to inhibitor, 298, **298**
  - cAMP activation of, 297, 297f
  - catalytic subunit, 298, 298f
  - chains, 403
  - consensus sequence, 295
  - defined, 286
  - gene stimulation, 403
  - in glycogen metabolism, 633, 633f
  - pseudosubstrate binding to, 298, 298f
  - regulation of, 297, 297f
  - in signal transduction, 415
- Protein kinase B (PKB), 809
- Protein kinase C (PKC), 810
- Protein kinase inhibitors, as anticancer drugs, 417
- Protein kinases
  - catalysis, 286
  - dedicated, 295
  - defined, 294
  - kinase fold, 298
  - multifunctional, 295
  - in phosphorylation, 294–296, 295t
  - serine, 294–295, 295t
  - in signal transduction, 415
  - specificity, 295
  - threonine, 294–295, 295t
- Protein phosphatase 1 (PP1)
  - functions of, 629, 633–635
  - in glycogen metabolism, 629, 633–635, 634f, 636
  - phosphorylase *a* and *b* and, 636
  - regulation in muscle, 634, 634f
  - regulation of glycogen synthesis by, 633–634, 634f
  - substrate for, 636
  - subunits, 634
- Protein purification, 78
  - affinity chromatography in, 70–71, 70f
  - assays in, 67
  - binding affinity and, 68–71
  - cell release and, 67–68
  - charge and, 68–71
  - dialysis in, 69, 69f
  - electrophoretic analysis of, 75f
  - gel electrophoresis in, 71–73, 72f, 73f
  - gel-filtration chromatography in, 69, 69f
  - high-performance liquid chromatography (HPLC) in, 71, 71f
  - homogenization in, 67
  - ion-exchange chromatography in, 69–70, 70f
  - isoelectric focusing in, 73, 73f
  - in protein studies, 65
  - quantitative evaluation of, 75–76, 75f
  - recombinant DNA technology in, 78–79
  - salting out in, 68
  - SDS-PAGE in, 73, 73f, 74
  - sedimentation coefficient and, 76, 77f
  - sedimentation-equilibrium technique in, 78
  - separation of proteins in, 71–75
  - size and, 68–71
  - solubility and, 68–71
  - two-dimensional electrophoresis in, 74–75, 74f
  - ultracentrifugation and, 76–78, 77f
  - zonal centrifugation in, 77, 77f
- Protein sorting
  - defined, 915
  - pathways, 918, 918f
- Protein structures, 2f, 3
  - amino acid sequence as determinant, 49–59
  - amino acid sequences and, 37, 37f
  - complex assembly, 28, 28f
  - dictating function, 28f
  - elucidation of, 100
  - family of, 100, 100f
  - models of, 22–23, 23f, 61–62
  - primary, 27, 35–40, 48, 78–79, 178–179
  - quaternary, 27, 48–49
  - secondary, 27, 40–46, 48
  - synthetic peptides and, 93
  - tertiary, 27, 46–48, 178–179
  - three-dimensional prediction from sequence, 54
- Protein studies
  - enzyme-linked immunosorbent assay (ELISA) in, 82–83, 83f
  - fluorescence microscopy in, 84, 84f
  - immunology in, 79–84
  - mass spectroscopy in, 85–92, 86f, 88f, 92f
  - peptide synthesis, 92–95
  - purification methods for, 66–79
  - steps in, 65–66
  - western blotting in, 83–84, 83f
  - x-ray crystallography in, 95–97
- Protein synthesis, 893–921. *See also* Translation
  - accuracy of, 894–895, 894t
  - adaptors in, 123, 123f
  - antibiotic inhibitors of, 913–915, 914f, 914t, 915f
  - endoplasmic reticulum (ER) in, 915–919
  - eukaryotic versus bacterial, 911–913
  - insulin in, 816
  - mRNA in, 126
  - peptide-ligation methods, 95
  - polypeptide-chain growth in, 894, 894f
  - ribosomes as site of, 893, 902–911
- Protein targeting, 915–919
- Protein turnover. *See also* Amino acid degradation
  - defined, 682
  - regulation of, 683–687
- Proteins, 27–63
  - acetylation of, 293–294
  - adaptor, 410
  - affinity tags and, 78
  - aggregated, in neurological diseases, 56–57, 56f
  - aggregation of, 683
  - allosteric, 285, 286–292
  - alpha helix of, 40–41
  - amino acids. *See* amino acids
  - antibody generation to, 79–80
  - β pleated sheets of, 44–45, 44f, 45f
  - as building block, 20
  - building blocks, 2
  - carbohydrate units of, 58
  - cargo, 918–919
  - cellular, degradation of, 682–683
  - chaperone, 50
  - cleavage of, 57–59, 88–89, 89f
  - coat (COPs), 919
  - coiled-coil, 44–45, 44f, **44,** 45f
  - covalent modification of, 293–298, 294t
  - crystallization of, 95
  - defined, 2, 36
  - degradation of. *See* amino acid degradation
  - denatured, 52, 52f
  - dietary, digestion and absorption of, 682, 683f
  - disulfide bonds, 36–37
  - DNA-binding, 926–932
  - enzymes as, 215
  - evolution of, 128
  - fibrous, 44–46
  - flexibility of, 29, 29f
  - functional groups of, 27–28
  - functions of, 27–29
  - genetically engineered, 152
  - glycan-binding, 334
  - glycosylation of, 327–332
  - half-life of, 683, 684, 685t
  - hemoglobin as model of, 191–212
  - interaction of, 28
  - intrinsically unstructured, 55–56
  - as linear polymers, 27–28, 35
  - loops, 44, 44f
  - membrane. *See* membrane proteins
  - metamorphic, 55–56
  - misfolding of, 56–57
  - molecular weights of, 76t
  - overview of, 65–66
  - plasma-membrane, 916, 916f
  - properties of, 27–29
  - pure, 66
  - refolding of, 50, 50f
  - regulatory, 930
  - release factors, 126
  - repressor, 928
  - reverse turns, 44, 44f
  - ribosomal. *See* ribosomes
  - rigidity of, 29
  - S values of, 76t
  - secretory, 916, 916f
  - secretory pathway, 915–916
  - separation of, 65, 71–75
  - subunits of, 48–49
  - tagging for destruction, 683–685
  - translocation of, 915–919
  - unfolding of, 52, 52f
  - unstructured, 55–56

Proteoglycans
  in cartilage, 328–329, 329f
  defined, 326
  properties of, 327–328
  structural roles, 327–328
Proteolysis, 216, 819
Proteolytic activation, 286, 299–308
  apoptosis by, 299
  of chymotrypsinogen, 299–300, 300f
  zymogens, 301, 301f
Proteolytic enzymes
  activation of, 299–308
  in blood clotting, 303–304
  chymotrypsinogen as, 299–301, 300f
  in digestion, 299–303
  function of, 216
  inhibitors of, 302–303
Proteomes, 66
Prothrombin, 304, 304f
Proton abstraction, 268
Proton gradients
  across thylakoid membrane, 578–581
  in ATP synthesis, 541–549
  cytochrome *bf* contribution to, 575, 575f
  directionality of, 575, 575f
  in oxidative phosphorylation, 433, 433f, 523, 524, 529, 541–549
  power transmission by, 558, 558f
Proton shuttle, 267–268, 268f
Proton transport, 538, 538f
Proton-motive force
  ATP forms without, 544–545, 545f
  in oxidative phosphorylation, 523, 542–543, 542f
  in photosynthesis, 578
Proto-oncogenes, 416
Protoporphyrin, 192
Protoporphyrin IX, 736, 737f
Proviral DNA, 159
Proximal histidine, 193, 200
PrP, 56
Prusoff, William, 1035
Pseudo-first-order reactions, 226
Pseudogenes, 154
Pseudosubstrate sequence, 297
Psoralens, 847, 883
P-type ATPases
  defined, 367, 370
  as evolutionarily conserved, 374
  SERCA, 370–373
Puffer fish, genome of, 157, 157f
Pumps
  action, 371, 371f
  ATP-driven, 367
  calcium ion, 371, 371f, 372–374, 372f
  defined, 367
  ion gradients, 368
  MDR, 374
  Na^+-K^+, 370, 373
  purification of, 370
Pur repressor, 931, 931f
Purine biosynthetic pathway, 696
Purine catabolism, 760f
Purine nucleotide synthesis, 24, 34, 724
  AMP in, 751–752
  control of, 758–759, 758f
  de novo pathway for, 450f, 748, 748f, 749t
  enzymes of, 752
  GMP in, 751–752
  phosphorylation in, 749–751, 749t, 750f
  ribose phosphate in, 749
  salvage pathway for, 748, 752
Purine nucleotides, **107,** 748
Purinosomes, 752, 752f
Puromycin, 914, 914t, **914**
Pyran, 318
Pyranose, 318, 319f, 320–321
Pyridoxal phosphate enzymes
  bond cleavage in, 691, 691f
  reaction choice, 691–692, 692f
  stereoelectronic effects, 691, 691f
Pyridoxal phosphate (PLP), 621, 621f
  in amino acid degradation, 682, 691–692
  in Schiff-base intermediates formation, 689–690
  in Schiff-base linkage, 720
Pyridoxamine phosphate (PMP), 690, 720, 720f
Pyridoxine (vitamin B_6), 689
Pyrimidine biosynthetic pathway, 696
Pyrimidine nucleotide synthesis, 734
  carbamoyl phosphate in, 745, **745**
  control of, 758
  cytidine triphosphate (CTP) in, 747–748
  de novo pathway for, 744, 744f
  glutamine in, 745
  mono-, di-, and triphosphates in, 747
  orotate in, 746–747, 746f
  substrate channeling in, 745–746, 746f
Pyrimidine nucleotides
  defined, 744
  recycling of, 748
Pyrimidines, **107**
Pyrosequencing, 155–156
Pyrrolidine rings, 45, 45f
Pyruvate, **232**
  in amino acid degradation, 692–693, 698, 699–700, 699f
  in ATP formation, 460–461
  carboxylation of, 479–480, 514
  conversion into phosphoenolpyruvate, 478–480, 479f
  fates of, 462, 462f
  glucose conversion into, 461–462
  in glycolysis, 452f
  metabolism disruption, 515–516
  oxaloacetate synthesis from, 440–441, 440f
  reduction of, 464, 464f
Pyruvate carboxylase
  in citric acid cycle, 514–515
  in gluconeogenesis, 478–480, 479f
Pyruvate decarboxylase, 462–463, 513
Pyruvate dehydrogenase, in citric acid cycle, 511–512
Pyruvate dehydrogenase complex
  carbon dioxide production, 498–500
  in citric acid cycle, 497–501, 497f, 497t, 501f, 511–512, 511f
  components of, 498–500, 499f
  of *E. coli*, 497f, 497t
  mechanism, 498–500, 499f
  reactions of, 500–501, 501f
  regulation of, 511–512, 511f
  response to energy change, 511, 511f
Pyruvate dehydrogenase kinase (PDK), in cancer, 513
Pyruvate dehydrogenase phosphatase, 511
Pyruvate kinase
  liver and glycolysis and, 473, 473f
  muscle and glycolysis and, 471
Pyruvate kinase M, 475

Q cycle
  defined, 535
  in respiratory chain, 535, 535f
Q-cytochrome *c* oxidoreductase
  defined, 533–534
  in respiratory chain, 529–531, 530f, 530t, 533–534
  structure of, 534, 534f
Quantitative PCR (qPCR), 157–158, 158f
Quaternary protein structure, 27, 48f
  complex, 49f
  defined, 48
Quinones
  oxidative states of, 531, 531f
  ubiquitous, 530
Quinonoid intermediate, 690
Quorum sensing, 934, 934f

R groups. *See* Amino acid side chains
R state, of hemoglobin, 198, 199f, 200, 211–212
Racker, Efraim, 542b
Raf, 414
Raloxifene, **952**
Ramachandran plot
  for angles of rotation, 40f
  β stands, 42f
  defined, 39
  for helices, 40f
Random-coil conformation, 50
Ranitidine (Zantac), 1049, **1049**
Ras
  activation of, 413, 413f
  affixation to cytoplasmic face, 294
  GTPase activity of, 414, 414t
  in protein kinase cascade initiation, 414
Rate constants, 225–226
RBP4, 806
Reactants, 219
Reaction centers
  accessory pigments and, 581–584, 583f
  bacterial, 569–572, 570f, 571f
  cyclic electron flow in, 572, 572f
  defined, 569

  - photosystem I, 572, 576, 576f
  - photosystem II, 572, 573
- Reaction rates
  - enzymes and, 220–221, 220f
  - free-energy difference and, 218
  - kinetics as study of, 225–226
  - rate constants, 225–226
- Reactions
  - acid-base, 13–15
  - ATP hydrolysis, 279, 279f
  - bimolecular, 226
  - in Calvin cycle, 590–597, 597f
  - carbonic anhydrases and, 264–268
  - of citric acid cycle, 508
  - deleterious, 456
  - DNA ligase, 832, 832f
  - double-displacement, 232, 233
  - enzyme acceleration of, 221–225
  - equilibrium, 220–221
  - equilibrium constant of, 220
  - in fatty acid oxidation, 651t
  - in fatty acid synthesis, 662–664
  - first-order, 225
  - free-energy difference of, 218, 219
  - of gluconeogenesis, 482t
  - group-transfer, 441, 441f
  - hydrolytic, 441–442, 442f
  - isomerization, 441, 441f
  - ligation, 440–441, 440f
  - multiple-substrate, 231–233
  - oxidation-reduction, 440, 440f
  - oxygenase, 593–594, 593f
  - pentose phosphate pathway, 605–607, 605f, 606f
  - polymerization, 1018
  - pseudo-first-order, 226
  - of pyruvate dehydrogenase complex, 500–501, 501f
  - reduction potentials of, 527t, 528–529
  - second-order, 226
  - sequential, 232
  - standard reduction potentials of, 527
  - transaldolase, 606–607, 606f
  - transketolase, 605–606, 605f
  - velocity versus substrate concentration, 222, 222f
- Reactive oxygen species (ROS)
  - defined, 539
  - glucose 6-phosphate dehydrogenase in protection against, 610–612
  - pathological conditions that may entail injury, 539t
  - release prevention, 194–195
  - source of, 656
- Reactive substrate analogs, 237–238, 237f, 238f
- RecA, 853, 933
- Receptor-mediated endocytosis, 360–361, 360f, 784, 784f
- *Reclinomonas americana*, 526
- Recognition helix, 927
- Recognition sites (sequences), 269
  - in cognate and noncognate DNA, 272–274, 273f, 274f
  - defined, 269
  - distortion of, 273–274, 273f
  - EcoRV endonuclease, 272, 272f
  - structure of, 272, 272t
  - tRNA template, 123
- Recombinant DNA
  - formation of, 143–144
  - manipulation of, 143
- Recombinant DNA technology, 135–166
  - amino acid sequence information, 89–90
  - bacterial artificial chromosomes in, 147
  - in biology, 143–152
  - blotting techniques in, 136, 138, 138f
  - cloning in, 147–148
  - cohesive-end method in, 144, 144f
  - complementary DNA (cDNA) in, 149–150, 149f, 150f
  - defined, 135
  - in disease-causing mutations identification, 143
  - DNA probes in, 138, 139–141
  - DNA sequencing in, 136, 138–139
  - DNA synthesis in, 139–141, 140f, 141
  - electroporation in, 163, 164f
  - functional effects of disease-causing mutations and, 152
  - λ phase in, 144–146, 147f
  - gel electrophoresis in, 137–138, 138f
  - gene disruption in, 160–163, 160f, 161f
  - gene expression analysis in, 157–159, 158f, 159f
  - gene therapy in, 164
  - genomic libraries in, 147–148, 148f
  - mutation creation, 150–152
  - in mutation identification, 143, 143f
  - overview of, 135–136
  - in plants, 163–164, 163f
  - plasmids in, 144–146
  - polymerase chain reaction (PCR) in, 136, 141–143, 141f, 142
  - in protein purification, 78–79
  - requirement, 136
  - restriction-enzyme analysis in, 136, 137
  - RNA interference in, 162–163, 162f
  - solid-phase approach in, 139–141, 140f
  - tools of, 136–143
  - transgenic animals in, 159f, 160
  - vectors in, 145–146
  - yeast artificial chromosomes (YACs) in, 147, 147f
- Recombinases, 854
- Recombination
  - in antibodies, 991–995, 991f, 992f
  - in color blindness, 974–975, 974f
  - DNA. *See* DNA recombination
  - homologous, 974b
- Recombination signal sequences (RSSs), 992
- Recombination synapse, 854
- Redox balance, 463, 463f
- Redox couples, 527
- Redox potential. *See* Reduction potential
- Reducing sugars, 321–322
- Reductase mRNA, in cholesterol regulation, 781
- 5α-reductase, 793
- Reduction potential
  - measurement of, 526, 526f
  - of NAD^+, 528
  - in oxidative phosphorylation, 526–528, 526f
  - of reactions, 527t, 528–529
- Reed, Randall, 963
- Regulator genes, 928
- Regulatory domains
  - in amino acid synthesis, 731–732
  - recurring, 731f
  - structures of, 732f
- Regulatory light chain, 1012
- Regulatory protein P_{II}, 733, 733f
- Regulatory proteins, 930
- Regulatory sites, 287
- Relay helix, 1016, 1016f
- Release factors
  - defined, 126
  - ribosome (RRF), 911
  - translation termination by, 910–911, 911f
- Repeating motifs, 91, 91f, 180–181, 180f
- Replication factor C (RFC), 844
- Replication fork. *See also* DNA replication
  - defined, 831
  - schematic view of, 840f
- Replication protein A, 843
- Replicon, 843
- Reporter genes, 145
- Repressors
  - corepressors, 931
  - λ, 932–934, 933f
  - *lac*, 928–930, 929f, 930f
  - protein, 928
  - *pur*, 931, 931f
- Resistin, 806
- Resonance energy transfer, 582–583, 582f
- Resonance structures
  - ATP (adenosine triphosphate), 429
  - defined, 7
  - depiction of, 7
  - improbable, 429f
- Respirasome, 529
- Respiration, 524b
- Respiratory chain
  - coenzyme Q in, 530–531
  - components of, 530f, 530t
  - cytochrome *c* oxidase in, 529, 530f, 530t, 535–538
  - defined, 523
  - electron transfer in, 540–541, 540f
  - NADH-Q oxidoreductase in, 529–531, 530f, 530t, 532–533, 533f
  - in oxidative phosphorylation, 529–541
  - Q cycle in, 535, 535f
  - Q-cytochrome *c* oxidoreductase in, 529, 530f, 530t, 533–534
  - succinate-Q reductase in, 529, 530f, 530t, 533

Respiratory control, 553–554, 554f
Respiratory distress syndrome, 774
Respiratory quotient (RQ), 815
Restriction enzymes (endonucleases), 251
  analysis, 137
  defined, 137b, 269
  in DNA cleavage, 137, 269–275
  in *E. coli*, 269
  in forming recombinant DNA molecules, 143–144
  hydrolysis of phosphodiester bond, 269–270, 269f
  inverted repeats, 272, 272f
  recognition sites, 272, 272f, 273–274, 273f
  restriction-modification systems, 274
  specificity, 137, 137f, 272, 274
  type II, 275
Restriction fragments
  separation by gel electrophoresis, 137–138, 138f
  in Southern blotting, 138
Restriction-modification systems, 274
11-*cis*-retinal, 970, 971–972, 972f
Retinitis pigmentosa, 884–885
Retrovir, 263
Retroviruses
  defined, 118
  in drug resistance evolution, 1053
  flow of information in, 119, 119f
  in gene introduction, 159
Reverse cholesterol transport, 787
Reverse transcriptase, 119, 119f
Reverse turns
  amino acid residues in, 51–52
  defined, 44
Reversible covalent modification, 286, 443, 732–733, 733f
Reversible inhibition, 234, 234f. *See also* Enzyme inhibition
Reversible terminator method, 155
Rhizobium bacteria, 714
Rho, 868, 868f
Rhodopseudomonas viridis, 569, 570f, 584
Rhodopsin, 400, 400f, 970–971, 971f
Rhodopsin kinase, 972
Rhodospirillum rubrum, 541
Ribbon diagrams, 62, 62f
Ribonuclease
  amino acid sequences in, 49–50, 49f
  denatured, 50
  reduction and denaturation of, 50f
  sequence comparison of, 169, **170**
  structure of, 49–51, 49f
Ribonuclease III (RNase III), 870
Ribonuclease P (RNase P), 870
Ribonucleic acid. *See* RNA
Ribonucleotide reductase
  as cancer therapy target, 759–760
  defined, 753
  in deoxyribonucleotide synthesis, 753–755
  elements of, 753, 753f
  mechanism, 754, 754f
  regulation of, 759, 759f
  stable radicals, 755
  structure of, **753**
  tyrosyl radicals, 753, 753f
Ribose, 106, **106,** 317, 318f, 320f
Ribosomal initiator element (rInr), 873
Ribosome release factor (RRF), 911
Ribosomes
  bacterial, 911
  catalysis by proximity and orientation, 908
  defined, 126, 893
  endoplasmic reticulum bound, 915–919, 916f
  eukaryotic, 911
  50S subunit of, 902–903, 903f, 908, 910
  polysomal, 910, 910f
  in protein synthesis, 893, 902–911
  schematic representation, 905f
  70S subunit of, 903, 903f
  sites of, 905, 905f
  structure of, 903, 903f
  subunits of, 902–903, 903f
  30S subunit of, 902–903, 903f, 907, 907f
  in translation, 898–911
  tRNA-binding sites, 905, 905f
Riboswitches, 867, 867f
Ribulose 1,5-bisphosphate, 591–592, 591f, 592f, **592,** 595–597, 596f, 597f
Ribulose 5-phosphate
  in Calvin cycle, 602, 603f
  in pentose phosphate pathway, 603, 604, 608–609
Ricin, 915, 915b
Ricinus communis, 915, 915f
Rickets, 795
Rickettsia prowazekii, 525–526
Rieske center, 534
Rifampicin, 869, 869f
Ritonavir, 792
RNA. *See also* Transcription
  ATP binding of, 186–187, 186f, 187f
  backbone of, 106, 106f
  base composition of, 121, 121t
  defined, 107
  double-stranded, 162, 162f
  in *E. coli*, 119, 119t
  flow of information from DNA to, 119, 119f
  in gene expression, 119–120
  guide strand, 162–163
  hybridization with, 135
  as linear polymer, 105, 106
  messenger. *See* mRNA
  micro, 879, 879f, 956–957, 956f
  nucleotides, 107–108
  passenger strand, 162
  primer, 841
  ribosomal. *See* rRNA
  small interference (siRNA), 162, 162f
  small nuclear (snRNA), 882–883
  small regulatory, 879, 879f
  splicing of, 880–883, 880f, 881f, 882t
  stem-loop structure, 113, 113f
  structure of, 113–114, 113f, 114f, 187, 187f
  synthesis of, 120–121
  transfer. *See* tRNA
RNA editing, 879–880, 879f
RNA interference, 162–163, 162f
RNA polymerase holoenzyme complex, 865, **865**
RNA polymerase I
  in RNA synthesis, 872, 873, 873f
  rRNA production, 877
RNA polymerase II
  promoter region, 874
  in RNA synthesis, 872, 873, 873f
RNA polymerase III, 873–874, 873f, 877, 878f
RNA polymerases, 120–121, 873f
  activated precursors, 121
  active site of, 861
  backtracking and, 863–864, 863f
  catalytic action of, 861–871
  components of, 121
  defined, 859, 860
  divalent metal ion, 121
  in DNA replication, 831, 831f
  DNA templates and, 121–122, 121t, 861
  in *E. coli*, 123, 123f, 861–862, 862t
  eukaryotic, 860, **860,** 872–874, 872t, 873f
  eukaryotic promoter elements, 873, 873f
  functions of, 860–861
  illustrated, 120f
  prokaryotic, 860, **860**
  promoter sites in, 864–865
  in proofreading, 863
  RNA-directed, 118
  structures of, 860, **860**
  subunits of, 862, 862t, 865
  synthesis of, 121f
  in transcription, 861–871
  transcription mechanism, 121, 121f
RNA processing
  eukaryotic pre-rRNA, 877, 877f
  of pre-mRNA, 877f, 878–879, 878f
  in transcription of eukaryotes, 872, 883–884, 884f
  tRNA, 877, 878t
RNA sequences, comparison of, 182–183, 182f
RNA splicing. *See* Splicing
RNA-DNA hybrid
  backtracking, 863–864, 863f
  during elongation, 869
  lengths of, 863
  separation, 863, 863f
  translocation, 863, 863f
RNA-induced silencing complex (RISC), 162–163, 162f, 956
Rods, 970–971, 970f
Rofecoxib (Vioxx), 1048, **1048,** 1052
Rossmann, Michael, 465
Rossmann fold, 465
Rosuvastatin, 1044, **1044**

Rotational catalysis, in ATP synthase, 546, 546f
Rotenone, 556
rRNA
  in base pairing, 903
  defined, 120
  folding of, 903, 904f
  in structural scaffolding, 903
  transcription of, 870–871, 870f, 877, 877f
  in translation, 903–905, 904f
  types of, 120
Rubisco
  in Calvin cycle, 591–594, 591f, 592f, 593f, 598–599
  defined, 591
  magnesium ion in, 592, 592f
  oxygenase reaction, 593–594, 593f
  structure of, 591, 591f
Rubisco activase, 593

S_1 pockets, 258, 259f
Saccharomyces cerevisiae, 153, 941
Sakmann, Bert, 379
Salicin, 1043
Salicylic acid, 1043, **1043**
Salmonella test, 852, 852f
Salt bridges, 203, 203f
Salting out, 68
Salty taste, 966, 969
Salvage pathways. *See also* Nucleotide synthesis
  defined, 744
  for IMP and GMP synthesis, 762
  purine, 748, 752
  pyrimidine, 748
Sample half-cell, 526–527
Sandwich ELISA, 82–83, 83f
Sanger, Frederick, 37, 138, 152
Sanger dideoxy method, 138–139, 139f
Sarcomere, 1020, 1020f
Sarcoplasmic reticulum Ca^{2+} ATPase. *See* SERCA
Saturated fatty acids, 661
Schiff bases, 451, 606, 621
  in amino acid degradation, 689–690
  in amino acid synthesis, 717
  in aminoacrylate, 729, **729**
  defined, 621b
  in retinal, 971, 971f
Scrapie, 56
Screening libraries, 1044–1046, 1046f
Screw sense, 40–41, 40b, 40f
Scurvy, 821
SDS (sodium dodecyl sulfate), 72
SDS-PAGE, 73, 73f, 74
SDS-polyacrylamide gel electrophoresis, 350, 351
Second Law of Thermodynamics, 11–12, 13
Second messengers, 398, **399,** 406, 415
Secondary active transport, 367
Secondary antibody, 84
Secondary protein structure, 27, 40–46, 48, 51t
Secondary transporters, 376, 376f, 377
Second-order reactions, 226
Secretory pathway, 915–916
Secretory proteins, 916, 916f
Sedimentation coefficients, 76, 77f
Sedimentation equilibrium, 78
Sedoheptulose 7-phosphate, 603, 603f, 604f
Segmental flexibility, 987, 987f
Selectins, 335
Selective estrogen receptor modulators (SERMs), 952
Selectivity filter, 381, 381f, 382, 382f
Self-splicing
  defined, 886
  example of, 886, 886f
  intron structure, 887f
  mechanism, 887, 887f
Self-tolerance, 1005
Sensory systems, 961–979
  brain connection, 962, 962f
  hearing, 975–977
  olfaction, 962–966
  overview of, 961–962
  taste, 966–970
  touch, 977–978
  vision, 970–975
Sequence alignment
  with conservative substitution, 176f
  with gap insertion, 172–173, 173f
  of hemoglobin, 176f
  of identities only versus Blosum-62, 176f
  of leghemoglobin, 176f
  of repeated motifs, 180, 180f
  scoring system, 176, 176f
  shuffling and, 173–174, 173f
  statistical analysis of, 171–177
  statistical comparison of, 173, 173f
Sequence comparison methods, 169–170
Sequence identities, 172
Sequence template, 180
Sequential model
  allosteric enzymes, 290–291
  of hemoglobin-oxygen binding, 199, 199f
Sequential reactions, 232
SERCA
  defined, 370
  P-type ATPase, 370–373
  pumping by, 372, 372f
  structure of, 372
Serine
  in catalytic triads, 255–258
  in chymotrypsin, 253–254, 254f
  in cysteine synthesis, 722–723, 726
  defined, 32–33
  from glycine, 700
  in glycine synthesis, 722–723
  pyruvate formation from, 699f, 700
  in sphingolipid synthesis, 772, **772**
  sphingosine from, 734
  structure of, 32f
  synthesis of, 722, 722f, 726
Serine dehydratase, 692
Serine hydroxymethyltransferase, 723
Serine kinases, 294–295, 295t
Serine protease inhibitors, 302
Serine proteases
  catalytic strategies of, 251
  convergent evolution and, 181–182
Serotonin, synthesis of, 734, 734f
Serpins, 302
7TM (seven-transmembrane-helix) receptors
  biological functions mediated by, 400t
  defined, 399
  in drug development, 1048
  in glycogen metabolism, 627–629, 628f
  ligand binding to, 400–402
  in olfaction, 963
  oligomeric, 969
  phosphoinositide cascade activation, 404–405, 405f
  in signal termination, 404f
  in taste, 967–969, 967f
  in vision, 970–974
70S initiation complex, 907, 907f
70S subunit, of ribosomes, 903, 903f, 907
Severe acute respiratory syndrome (SARS), 1050
Severe combined immunodeficiency (SCID), 164, 760–761
Shaker channel, 380, 384
Shaker gene, 380
Shape complementarity, 830, 830f, 831f
Shemin, David, 735b, 736
Shikimate, in amino acid synthesis, 727–729
Shine, John, 906
Shine-Dalgarno sequences, 126, 906
Shuffled sequences, 173–174, 173f
Sialic acids, 772
Sickle-cell anemia, 17–18, 205–206, 205f, 206f
Side chains
  amino acid. *See* amino acid side chains
  hydrophilic, 47
  hydrophobic, 47
  polypeptide chains, 36, 36f, 47
Sigmoid oxygen-binding curve, 196
Sigmoidal curve, 291, 291f
Sigmoidal kinetics, 287, 287f
Signal transduction, 397–420
  abnormalities in cancer, 416–417, 416f, 417f
  abnormalities in cholera, 417–418
  abnormalities in whooping cough, 417–418
  adaptor proteins in, 410
  β-adrenergic receptor in, 399–400, 400f
  calcium ions in, 405–407, 406f, 415
  calmodulin in, 407, 407f
  cAMP in, 402–403
  cross talk, 398–399
  defects, 415–418
  defined, 397
  element recurrence in, 415
  epidermal growth factor in, 397, 411–414
  epinephrine in, 399–407

Signal transduction—(*continued*)
  evolution of, 415
  function of, 397, 398f
  G proteins in, 400–403, 401f, 402f
  GTPases and, 414, 414t
  insulin in, 407–411
  molecular circuits, 398–399
  monoclonal antibodies as inhibitors, 416–417
  nucleotides in, 743
  in olfaction, 964, 964f
  overview of, 397–399
  primary messenger, 398
  principles of, 398–399, 398f
  protein kinases in, 415
  protein kinase A in, 403, 415
  Ras in, 413–414, 413f
  second messengers, 398, 406, 415
  secondary messengers, **399**
  7TM receptors, 399–402, 400t, 404f
  Src homology domains in, 410, 410f, 413, 415, 416, 416f
  in vision, 972–973, 972f
Signal-recognition-particle (SRP), 916–918, 917f
Sildenafil, 1042–1043
Simple diffusion, 368
Single-molecule studies, enzyme, 242–243, 242f
Single-stranded-binding protein (SSB), 840, 840f
Singulo method, 242
Site-directed mutagenesis, 150–151, 151f
Slack, C. Roger, 599
Sliding DNA clamp, 839–840, 839f
Sliding-filament model, 1020–1021, 1021f
Small G proteins, 413
Small interference RNA (siRNA), 162, 162f
Small nuclear ribonucleoprotein particles (snRNPs), 882–883, 882t
Smell. *See* Olfaction
SNARE proteins, 361, 362f
Social interactions, in gene expression, 934
SOD1 gene, 159
Sodium dodecyl sulfate (SDS)
  in gel electrophoresis, 72–73, 73f
  PAGE, 73, 73f
Sodium ion channels
  amiloride-sensitive, 969, 969f
  inactivation of, 384–385
  paddles, 388
  purification of, 379–380
  in taste, 969
Sodium ions, 378
Solid-phase DNA synthesis, 139–141, 140f
Solid-phase peptide synthesis, 93–94, 94f
Somatic mutation, antibody diversity and, 992–993
Sonicating, 349
Sos, 413, 413f
Sour taste, 966, 969–970
Southern, Edwin, 138
Southern blotting, 138, 138f, 160
Space-filling models, 22, 23f, 61, 61f, 195f
Special pair, 570–571
Specificity constant, 230–231, 231f
Sphingolipids
  ceramide and, 772, 772f
  defined, 772
  diversity and, 773–774
  gangliosides, 772–773, 774
  synthesis of, 772–774, 773f
Sphingomyelin
  defined, 345, 772
  structure of, **345**
Sphingosine, 344
  defined, 345
  structure of, **345**
  synthesis of, 734, 734f
Spina bifida, 762
Spleen tyrosine kinase (Syk), 993
Spliceosomes
  assembly and action, 882–883, 882f
  catalytic center, 883, 883f
  defined, 128, 881
Splicing
  alternative, 128, 129f, 877, 885–886, 885f, 885t
  branch points in, 880, 881f
  catalytic center, 883, 883f
  defined, 128, 880
  group I, 887, 888f
  group II, 887–888, 888f
  mechanism, 881, 881f
  mutations affecting, 880–881, 884–885, 884f
  pathways comparison, 888f
  self-splicing and, 886–888, 886f, 887f
  sites of, 880–881, 880f
  snRNPs in, 882–883, 882t
  spliceosomes in, 881, 882–883, 882f
  transesterification in, 881–882, 882f
Splicing factors, 882
Split genes, 128
Split-pool synthesis, 1045, 1045f
Spudich, James, 1019
Squalene
  cyclization of, 778–779, 779f
  synthesis of, 777–779, 778f
Squalene synthase, 778
Src
  affixation to cytoplasmic face, 294
  in cancer, 416
Src homology 2 (SH2) domain, 410, 410f, 415, 416, 416f
Src homology 3 (SH3) domain, 413, 415, 416, 416f
SREBP (sterol regulatory element binding protein), 779–781, 780f, 781f, 784
SREBP cleavage activating protein (SCAP), 780–781, 781f
SRP receptor, 917, 917f
Stacking forces, 111
Stahl, Franklin, 115, 116
Standard free-energy change, 219–220
Standard reference half-cell, 526, 527
Staphylococcus aureus, 239, 239f
Starch, 324, 597
Starvation, metabolic adaptations in, 818–819, 819t
Starved-fed cycle, 816–817
Statins, 788
Steady-state assumption
  defined, 227
  in enzyme kinetics, 226
Stearoyl CoA desaturase, 668, 668f
Steatorrhea, 646
Steitz, Tom, 829
Stem-loop structure, 113, 113f
Stereochemical renderings, 22
Stereochemistry
  of cleaved DNA, 270–271, 271f
  observation of, 270
  of proton addition, 721, 721f
Stereocilia, 975–976
Stereoelectronic control, 691f, 692
Stereoisomers
  carbohydrate, 317, 317f
  notation for, 29b
Steric exclusion, 39
Steroid hormones
  anabolic, 951
  binding and activation, 789–790
  classes of, 789
  defined, 789
  hydroxylation of, 790–791
  identification of, 790
  synthesis of, 789–791, 789f
Sterol regulatory element binding protein (SREBP), 779–781, 780f, 781f, 784
Sterol regulatory element (SRE), 779
Sticky ends, DNA, 144, 144f
Stoichiometry
  of gluconeogenesis, 481
  for light reactions, 581
  of palmitate synthesis, 666
Strand invasion, 853, 853f
Strand separation, 141, 141f
Streptomyces coelicolor, 865
Streptomyces lividans, 380
Streptomycin, 914, 914t, **914**
Stroma, 567
Stroma lamellae, 567–568
Strominger, Jack, 1042
Structural genes, 928
Structure-activity relationship (SAR), 1046
Structure-based drug development, 1046–1048, 1047f, 1048f
Substituted enzyme intermediates, 233
Substitution matrix, 174–176, 175f, 176f
Substitutions
  in amino acid sequences, 174–177, 175f, 176f
  conservative, 174, 175f, 176f
  nonconservative, 174, 175f
Substrate binding
  in catalysis, 225
  conformation selection, 224
  cooperative, 233

  induced-fit model, 224, 224f, 252
  lock-and-key model, 224, 224f
Substrate channeling
  in amino acid synthesis, 730, 730f
  defined, 510
  in pyrimidine synthesis, 745–746, 746f
Substrate cycles, 485, 485f
Substrate-induced cleft closing, 453
Substrate-level phosphorylation, 460
Substrates. *See also* Enzyme-substrate complex
  accessibility of, controlling, 444
  in biochemical reactions, 231–233
  chromogenic, 254, 255f
  concentration, 226, 226f, 228
  defined, 216
  enzyme interaction, 223
  homotropic effect on allosteric enzymes, 290
  multiple, in reactions, 231–233
  reciprocal relation, 759
  spectroscopic characteristics of, 223
Subtilisin, **182**
  catalytic triad of, 259, 259f
  oxyanion hole of, 259, 259f
  site-directed mutagenesis of, 260, 260f
  structure of, 181–182
Subunit vaccines, 1006
Succinate
  in glyoxylate cycle, 517, 517f
  oxidation of, 507–508
Succinate dehydrogenase, in citric acid cycle, 507–508
Succinate-Q reductase, 529–531, 530f, 530t, 533
Succinyl CoA
  in amino acid degradation, 698, 701, 701f
  in citric acid cycle, 505–506, 506f
  formation of, 505, 654, 654f, 655–656, 656f
  porphyrins from, 736–737, 737f
Succinyl CoA synthetase, **507**
  $\alpha_2\beta_2$ heterodimer, 507
  in biochemical transformation, 506–507
  reaction mechanism of, 506, 506f
  structure of, 507, **507**
  subunits, 507, 507f
Sucrase, 323, 450
Sucrose
  defined, 324
  structure of, 323–324, 324f
  synthesis of, 598, 598f
Sucrose 6-phosphate, 598, 598f
Sugars
  in cyclic forms, 318–320
  DNA, 4
  five-carbon, 595, 596f, 601–607
  monosaccharides, 316–323
  N-linked, 330
  O-linked, 330
  phosphorylation, 323
  reducing, 321–322
Suicide inhibition, 238, 238f, 757, 761
Sulfhydryl groups, 33
Sulfolipids, 567
Super compensation (carbo-loading), 815
Supercoiling, 113
  ATP hydrolysis in, 837–839, 838f
  catalyzation of, 838
  defined, 834, 835
  degree of, 835–836
  DNA condensation from, 836
  linking number and, 834–836, 834f
  negative, 836, 836f, 837–839
  positive, 836
  relaxation of, 836–837, 837f
  right-handed, 835
  topoisomerases and, 836–839, 837f, 838f, **838**
  twist in, 834f, 835
  writhe in, 834f, 835–836
Superhelical cable, 45
Superoxide anion, 194
Superoxide dismutase, 539–540, 540f
Superoxide radicals, 538–540
Supersecondary structures
  defined, 48
  helix-turn-helix, 48f
Suppressors of cytokine signaling (SOCS), 806–807, 807f
Svedberg units, 76
Sweet taste, 966, 968–969
Symmetry matching, 927–928
Symporters, 376, 376f, 377
Synaptic cleft, 385, 386f
Synchrotron radiation, 95
Synonyms, 125
Synthase, 502b
Synthetic analog system model, 266–267, 267f
Synthetic peptides
  as antigens, 92
  construction of, 93–94, 94f
  as drugs, 92–93, 93f
  linking of, 95
  in receptor isolation, 92
  in three-dimensional structure of proteins, 93
Synthetic vaccines, 152

T cells
  activation of, 999–1000, 999f
  defined, 984
  helper, 984, 1000–1002
  HIV infection and, 1003–1004
  killer, 984, 998–1000
  memory, 1006
  negative selection in, 1004–1005, 1004f
  positive selection in, 1004, 1004f
T state, of hemoglobin, 198, 199f, 200, 203, 211–212
T1R proteins, as sweet receptors, 968–969, 968f
T2R proteins, as bitter receptors, 967–968, 968f
Tamoxifen, **952**, 952f
Tandem mass spectroscopy, 87–88, 88f
TAP proteins, 996
Taste, 966–970. *See also* Sensory systems
  anatomic structures in, 967, 967f
  bitter, 966, 967–968
  ion channels in, 969, 969f
  salty, 966, 969
  7TM receptors in, 967–969, 967f
  sour, 966, 969–970
  sweet, 966, 968–969
  tastants in, 966–967, 966f, **966**
  types of, 966
  umami, 966, 969
Taste buds, 967, 967f
TATA box
  defined, 122
  in transcription of eukaryotes, 874, 874f
TATA-box-binding protein (TBP), 180–181, 874–875, 875f, 953
Taxol, 1023–1024
Tay-Sachs disease, 774
T-cell receptors (TCRs)
  CD3 and, 999, 999f
  CD4 and, 1001, 1001f, 1004–1005
  CD8 and, 998–1000, 999f, 1004–1005
  CD28 and, 1000
  CD45 and, 1000
  defined, 982, 998
  docking mode, 985
  domains of myosin, 998, 998f
  genetic diversity of, 985, 998
  helper T cells and, 1000–1002
  killer T cells and, 998–1000
  MHC proteins and, 998–1003, 998f
  structure of, 998f
  in T-cell activation, 999–1000, 999f
Telomerase, 845, 845f
Telomeres
  defined, 844
  in DNA replication, 844–845, 845f
Temperature, 11–12
Template
  coding strands and, 862, 862f
  complementarity with mRNA, 122, 122f
  defined, 829b
  DNA polymerases, 829
  lagging strand, 841
  in replication, 828–829
  RNA polymerases and, 121–122, 121t, 861
  transcribed regions of, 866–867
  transcription bubbles and, 866
  transcription in, 123
Ternary complex, 232
Tertiary protein structure, 27, 46–48, 178–179, 179f
Testosterone, 789, 793, **793**
Tetrahydrobiopterin, 703, **703**
Tetrahydrofolate, 723f
  in amino acid synthesis, 723–724
  defined, 723
  one-carbon groups carried by, 723t
Tetrahymena, 886, 887
Tetrapyrrole, 736
Tetraubiquitin, **684**

Tetroses, 316
Thalassemia, 207, 884, 884f
Therapeutic index, 1041
Thermodynamics
  of coupled reactions, 425–426, 427–429
  laws of, 10–12
  of metabolism, 425–426
Thermogenin, 555
Thiamine (vitamin B_1), 438t, 515
Thiamine pyrophosphate (TPP)
  in citric acid cycle, 499, 499f
  deficiency, 515
  defined, 515
Thick filaments, in myosin, 1020, 1021f
Thin filaments, in actin, 1020
Thioester intermediate, 457–459, 458f
Thioesterase, 664
6-thioguanine, 1051, **1051**
Thiol groups, 33
Thiopurine methyltransferase, 1051
Thioredoxin, 579–580
  in Calvin cycle, 599, **599**
  enzyme activation by, 599f
  enzymes regulated by, 599t
Thioredoxin reductase, 754
30S initiation complex, 907, 907f
30S subunit, of ribosomes, 902–903, 903f, 907
-35 region, 122
Three-dimensional structure
  conservation of, 178–179, 179f
  in sequence alignment evaluation, 179–180
  tertiary structure and, 178–179
  in understanding evolutionary relationships, 177
Threonine
  defined, 33
  pyruvate and, 700
  structure of, 32f
  synthesis of, 732, 732f
Threonine deaminase, 731–732, 731f
Threonine dehydratase, 692
Threonine kinases, 294–295, 295t
Threonine operon, 936, 936f
Threonyl-tRNA synthetase, 899–900
Threonyl-tRNA synthetase complex, 901, **901**
Threshold effects
  allosteric enzymes, 291, 291f
  defined, 291
Thrombin, 217, 217f
  antithrombin and, 307–308
  in blood clotting, 304–305, 307
  dual function of, 307
  inhibitors of, 307
Thromboxane synthase, 669
Thromboxanes, 669
Thylakoid membranes
  defined, 567–568
  pH gradient across, 578
  photosynthesis in, 567–568
  photosystem II and, 574–575
  proton gradient across, 578–581
  stacked, 583
  unstacked, 583
Thylakoid spaces, 567–568, 578
Thylakoids, 567
Thymidine, 107, 748
Thymidine kinase, 748
Thymidine monophosphate (TMP), 757
Thymidine phosphorylase, 748
Thymidylate, 107
  blocking synthesis of, 757–758
  defined, 755
  synthesis of, 755–756, 756f
Thymidylate synthase, 756
Thymine, **4,** 5f, 110, **110,** 748
  in DNA repair, 850–851
  synthesis of, 724
Thymocytes, 1004
Thyroxine, synthesis of, 734, 734f
Time-of-flight (TOF) mass analyzer, 85
Tip links, 976, 976f
Tissue factor pathway inhibitor (TFPI), 307
Tissue factor (TF), 304
Tissue plasminogen activator (TPA), 129f, 308, 308f
Titin, 36
Toll-like receptors (TLRs)
  defined, 982
  extracellular domain of, 983, 983f
  illustrated, 982f
  PAMPs and, 982–983, 983t, 984f
Tonegawa, Susumu, 991
Topoisomerases
  bacterial, 839
  defined, 836
  type I, 836–837, 837f
  type II, 836, 837–839
Topoisomers, 835–836, 836f
Torpedo marmorata, 386, 386f
Torr, 196b
Torsion angles, 39, 39b
Tosyl-L-phenylalanine chloromethyl ketone (TPCK), 237–238, 237f, **237**
Touch, 977–978
Toxicity, drug, 1040–1041, 1049f
Toxins, 1006
Toxoid vaccines, 1006
Trace elements, 20
Trans configuration, peptide bonds, 38, 38t, 39t
Trans unsaturated fatty acids, 661
Transaldolase
  in pentose phosphate pathway, 602–605, 606–607, 606f
  reactions, 606–607, 606f
Transaminases
  in amino acid degradation, 688–691
  defined, 688
  mechanism, 690, 690f
Transamination
  in amino acid degradation, 689–691
  amino acid synthesis by, 720–721, 720f
  mechanism of action, 690, 690f
  Schiff base intermediates in, 689
Transcription, 859–890
  ATP hydrolysis in, 868, 868f
  in bacteria, 861–871
  of β-globin gene, 128, 128f
  chemistry of, 861
  defined, 859
  in DNA templates, 123
  elongation, 860, 862–863, 863f, 866, 866f
  in gene expression, 910
  inhibition, by antibiotics, 869–870, 869f
  mRNA processing and, 883–884, 884f
  nuclear hormone receptors in, 950–951, 951f
  overview of, 859–860
  promoter sites for, 122, 122f
  promoters in, 864–865, 864f
  proofreading in, 863
  regulation of, 871–876
  rho in, 868, 868f
  riboswitches in, 867, 867f
  RNA polymerases in, 121, 121f, 861–871
  of rRNA, 870–871, 870f, 877, 877f
  self-splicing in, 886–888, 887f, 888f
  splicing in, 860, 876, 880–883, 880f, 881f, 882t
  stages of, 860–861
  start site, 864
  termination of, 860, 866–867, 867f
  of tRNA, 870–871, 870f
  upstream promoter elements in, 864, 873
Transcription activator-like effector nucleases (TALENs), 161–162, 161f, 162f
Transcription bubbles
  defined, 862
  elongation at, 866, 866f
  schematic representation, 866f
  structure of, **862**
Transcription factors
  activation domain, 945, 946
  defined, 160, 945
  DNA-binding domains of, 945–946
  in eukaryotes, 874, 875–876, 945–947
  in gene expression, 945–947
  in induced pluripotent stem (iPS) cells, 947, 947f
  nuclear hormone receptor, 949–951, 950f
  in prokaryotes, 945
  regulatory domains of, 946
Transcription in eukaryotes, 871–886
  in bacteria versus, 871–872, 871f
  CAAT box in, 874, 874f
  carboxyl-terminal domain (CTD) in, 872
  downstream core promoter element (DPE) in, 874
  enhancers in, 876
  eukaryotic promoter elements, 873f
  GC box in, 874, 874f
  initiation of, 872, 874–875, 875f
  initiator element (Inr) in, 874
  microRNAs in, 879, 879f

  - nuclear membrane in, 871
  - pre-mRNA processing and, 877, 877f, 878–879, 878f, 884–885, 884f
  - products of, 876–886
  - RNA editing in, 879–880, 879f
  - RNA polymerases in, 872–874, 872t, 873f
  - RNA processing and, 872, 883–884, 884f
  - splicing in, 876, 880–883, 880f, 881f, 882t, 883f
  - TATA box in, 874, 874f
  - transcription factors in, 874, 875–876
  - translation and, 871–872, 871f
  - upstream promoter elements in, 873
- Transcription initiation, 860
  - in bacteria, 862–863
  - de novo, 862–863
  - in eukaryotes, 872, 874–875, 875f
- Transcription repressors, 946
- Transcriptome, 158
- Transducin
  - rhodopsin and, 972
  - α subunit of, **280**
- Transferases, 245t
- Transferred DNA (T-DNA), 163
- Transferrin, 361, 954
- Transferrin receptor, 361, 361f, 954
- Transferrin-receptor mRNA, 955, 955f
- Transgenic mice, 160
- Transglutaminase, 306
- Transition state
  - ATP hydrolysis, 277–278
  - in catalysis, 222, 241
  - catalytic stabilization of, 215
  - defined, 221
  - formation facilitation of, 221–225
  - symbol, 221
- Transition-state analogs, 234, 240–241, 241f, 242, 277, 277f
- Transketolase
  - in Calvin cycle, 595, 596f
  - in pentose phosphate pathway, 602–606, 605f
  - reaction, 605–606, 605f
- Translation, 893–921
  - accuracy of, 894–895, 894t
  - activation sites, 899–900, 900f
  - adenylation in, 899–902
  - aminoacyl-tRNA synthetase in, 894, 898–902
  - anticodons in, 896, 898, 898f, 901
  - in bacteria, 906, 907f, 911–913
  - base pairing in, 896, 903
  - codon-anticodon interactions in, 898
  - codons in, 897–898, 898f
  - defined, 893
  - direction of, 910
  - elongation factors in, 907–908, 907f, 913
  - in eukaryotes, 871–872, 871f, 911–913
  - eukaryotic initiation of, 911–912, 912f
  - formylmethionyl-tRNA in, 906, 906f, 907
  - in gene expression, 910
  - initiation factors, 907, 907f
  - initiation of, 905–906, 907f, 911–912
  - initiation sites, 906, 906f
  - mechanism of action, 909
  - mRNA in, 912
  - organization in, 913
  - overview of, 893–894
  - peptide bond formation in, 908–909, 908f
  - proofreading in, 900–901
  - reading frame, 907
  - release factors in, 910–911, 911f
  - ribosomes in, 898–911
  - rRNA in, 903–905, 904f
  - Shine-Dalgarno sequences in, 906
  - signaling in, 915–919
  - termination of, 910–911, 911f, 913
  - translocation in, 909–910, 910f
  - tRNA in, 893, 894, 895–898, 905, 905f, 911–912
  - wobble hypothesis and, 897, 897t
- Translesion, 846
- Translocase, 909
- Translocation
  - inhibition by antibiotics, 914–915, 915f
  - of proteins, 915–919
  - RNA-DNA hybrid, 863, 863f
  - signal sequences in, 916–918, 916f
  - in translation, 909–910, 910f
- Translocon, 917
- Transmembrane helices, 354–356, 355f, 355t
- Transmissible spongiform encephalopathies, 56–57
- Transplant rejection, 1002–1003
- Transport vesicles, 918
- Transporters
  - ABC, 367, 374–376, 374f, 375f, 1054
  - antiporters, 376, 376f
  - ATP-ADP, 551–552, 551f, 552f
  - dicarboxylate, 552, 552f
  - glucose, 368, 473–474, 474t
  - glucose 6-phosphate, 638
  - mitochondrial, 551–552, 551f, 552f
  - in oxidative phosphorylation, 549–552
  - phosphate, 552, 552f
  - in photosynthesis, 568–572, 569f
  - pyruvate, 552
  - secondary, 376–377
  - symporters, 376, 376f
  - tricarboxylate, 552, 552f
  - uniporters, 376, 376f
- Transverse diffusion, 357, 357f
- Tree of life, 3f
- Triacylglycerol synthesis, 768–776
  - liver in, 769
  - sources of intermediates in, 768, 768f
- Triacylglycerol synthetase complex, 768
- Triacylglycerols
  - in adipose tissue, 818
  - defined, 645
  - as energy source, 645, 645f
  - in fatty acid metabolism, 647–648, 647f
- Tricarboxylic acid (TCA) cycle. *See* Citric acid cycle
- Trimethoprim, 758
- Trimethylamine (TMA), 771
- Trimethylamine-*N*- oxide (TMAO), 771
- Trinucleotide repeats, 850
- Triose phosphate isomerase (TPI), 238, **238**
  - catalytic mechanism of, 455
  - in glycolysis, 455–456, 456f
  - structure of, 455f
- Trioses, 316
- Triple helix, of collagen, 45
- tRNA
  - acceptor stem, 896, 897
  - as adaptor molecules, 123, 124f
  - amino acid attachment to, 123, 123f
  - amino acid-attachment site, 123
  - anticodons of, 896, 897
  - bases, 896
  - CCA terminal region of, 896, 896f
  - charged, 898
  - codons in, 897–898
  - common features of, 895–897
  - defined, 120
  - extra arm, 896
  - helix stacking in, 896, 896f
  - initiator, 911
  - inosine in, 898
  - molecule design, 895–897, 895f
  - precursors of, 896
  - recognition sites on, 901, 901f
  - RNA processing, 877, 878t
  - structure of, 896, 896f
  - synthetase recognition of, 901, 901f
  - template-recognition site, 123
  - transcription of, 870–871, 870f
  - in translation, 893, 894, 895–898, 905, 905f
  - translocation of, 909–910, 910f
  - wobble hypothesis and, 897, 897t
- Trombone model, 841, 841f
- Tropomyosin, 1020, 1021
- Troponin complex, 1020
- TRP channels, 977
- *Trp* operon, 936
- Trypanosomes, 880
- Trypsin, 217, 217f
  - catalytic triad in, 258, 258f, 259f
  - chymotrypsin and, 258, 258f, 259f
  - defined, 301
  - generation of, 301
  - S_1 pockets of, 259f
  - structure of, 258, 258f
- Trypsin inhibitor, 302–303, 302f
- Trypsinogen, 301
- Tryptophan
  - in alanine, 700
  - degradation of, 704, 704f
  - nicotinamide from, 734, 734f
  - serotonin from, 734, 734f
  - structure of, 31f, 32
  - synthesis of, 728–729, 729f
- Tryptophan synthase, 729–730
- Ts elongation factor (EF-Ts), 907–908
- t-SNARE, 919

T-to-R equilibrium, 291–292, 292f
T-to-R state transition, ATCase, 290–291, 290f
Tu elongation factor (EF-Tu), 907, 907f, 908
Tuberculosis, 687
Tubulin, 1023, 1023f
Tumor growth, ceramide metabolism as stimulant, 774–775
Tumor hypoxia, 475
Tumor suppressor genes, 851
Tumor-inducing (Ti) plasmids, 163–164, 163f
Turnover number, 230, 230f
2′, 3′-dideoxy analog, 139
Two-dimensional electrophoresis in, 73
Twofold rotational symmetry, 272, 272f
Type 1 diabetes
  defined, 808
  glucagon excess in, 812
  insulin insufficiency in, 812
  metabolic derangements in, 812, 812f
  treatment for, 812
Type 2 diabetes
  defined, 808
  metabolic syndrome and, 809–810, 810f
  treatments for, 811–812
Type I topoisomerases, 836–837
Type II topoisomerases, 836, 837–839
Typing-monkey analogy, 53, 53f
Tyrosine, **703**
  defined, 32
  degradation of, 703, 704f
  structure of, 32f
  synthesis of, 727–728, 728f
  thyroxine from, 734, 734f
Tyrosine kinases, 294
Tyrosine phosphatase IB, 809
Tyrosinemia, 705t
Tyrosyl radical, 753, 753f
Tyrosyl-tRNA synthetase, 901

Ubiquinol, 533, 651
Ubiquinone, 530–531, 651
Ubiquinone reductase, 651
Ubiquitin
  conjugation, 683–684, 684f
  defined, 681, 683
  pathway, 686
  in protein tagging, 683–685
  structure of, 683–684
Ubiquitination, 684–686
UDP (uridine diphosphate), 632
UDP-galactose 4-epimerase, 467
UDP-glucose
  defined, 466, 630, 743
  in galactose conversion, 467
  glucose transfer from, 630–631
  synthesis of, 630
UDP-glucose pyrophosphorylase, 630
Ultracentrifugation, 76–78
Umami taste, 966, 969
UMP kinase, 747
Uncompetitive inhibition. *See also* Enzyme inhibition
  defined, 234
  double-reciprocal plot, 236, 236f
  example of, 235
  kinetics of, 235, 235f
Uncoupling proteins, 555–556, 555f
Unfolded protein response (UPR), 811
Uniporters, 376, 376f
Unsaturated fatty acids, 652–654, 652f, 653f, 661, 668–670
Unstructured proteins, 55–56
Upstream promoter elements, 864, 873
Uracil, 107, **107**
Uracil DNA glycosylase, 850, 850f
Urate, 761
Urea, 49, 694, 695b
Urea cycle
  in amino acid degradation, 693–698
  carbamoyl phosphate in, 693–695
  defined, 681
  disorders of, 697–698, 697f, 698f
  enzymes, 696–697
  in gluconeogenesis, 696, 696f
  illustrated, 693f
Ureotelic, 693
Ureotelic organisms, 693, 698
Uric acid, 698
Uridine, 107
Uridine monophosphate synthetase, 747
Uridine triphosphate (UTP), 286, 630
Uridylate, 108, 747
Uroporphyrinogen III, 736, 737, 737f

V genes, 991–992
Vaccines
  defined, 1006
  in disease prevention/eradication, 1006–1007
  HIV, 1007
  killed (inactivated), 1006
  live attenuated, 1006
  subunit, 1006
  synthetic, 152
  toxoid, 1006
Vaccinia virus, 160
Valine
  degradation of, 702–703
  in maple syrup urine disease, 705
  structure of, 31f, 32
van der Waals forces, 224
van der Waals interactions
  in antigen-antibody binding, 990
  base stacking, 111, 111f
  contact distance, 8
  defined, 8
  energy of, 8, 8f
  hydrocarbon tails, 348
  minimization of, 10
  as noncovalent bond, 7
Vanishing white matter (VWM) disease, 913, 913f
Variable number of tandem repeats (VNTR) region, 329, 329f
Vascular endothelial growth factor (VEGF), 187, 476
Vasopressin, 92–93, 93f
VDAC, 524
V(D)J recombination, 992, 992f
Vectors
  cloning, 145, 146, 147f
  defined, 143
  expression, 146
  plasmid, 144–146, 145f
  viral, 159–160
Very-low-density lipoproteins (VLDLs), 782, 782t
Vibrio fischeri, 934
Vicine, 611, 611f
Vioxx (rofecoxib), 1048, **1048,** 1052
Viral hepatitis, 691
Viral receptors, 335, 335f
Viral vectors, 159–160
Virions, 146
Viruses
  coats, 49
  in gene introduction, 159
  human immunodeficiency, 1003–1004, 1003f
  infectious mechanisms of, 335–336
  influenza, 335–336, 335f
  progeny, 146
  retroviruses, 118–119, 119f, 1053
  RNA, 118–119
Vision, 970–975. *See also* Sensory systems
  in animals, 974, 974f
  calcium ion in, 972–973
  color, 970, 973–975
  color blindness and, 974–975, 974f
  cones in, 970, 973–974
  evolution of, 974, 974f
  light absorption in, 971–972
  photoreceptors in, 970, 973–974, 973f
  retinal-lysine linkage in, 971, 971f
  rhodopsin in, 970–971, 971f
  rods in, 970–971, 970f
  7TM receptors in, 970–974
  signal transduction in, 972–973, 972f
Visual pigments
  in cones, 973–974
  defined, 970
  in rods, 970–971
Vitamin A (retinol), 439, 439f, 439t
Vitamin B_1 (thiamine), 438t, 515
Vitamin B_2 (riboflavin), 438t
Vitamin B_3 (niacin), 438t
Vitamin B_5 (pantothenic acid), 438t
Vitamin B_6 (pyridoxine), 438t
Vitamin B_7 (biotin), 438t
Vitamin B_9 (folic acid), 438t
Vitamin B_{12}, 438t
  as coenzyme, 654–655, 655f, 656
  corrin ring and cobalt atom, 654–655
  in fatty acid metabolism, 654–655

Vitamin C (ascorbic acid)
  deficiency of, alcohol related, 821–822
  forms of, 822, 822f
  function of, 439, 439t
Vitamin D (calcitriol)
  biochemical role, 795
  cholesterol as precursor, 794–795
  deficiency, 795
  function of, 439, 439t
  structure of, 439f
  synthesis of, 794–795, 794f
Vitamin E (α-tocopherol)
  function of, 439–440, 439t
  structure of, 439f
Vitamin K
  in blood clotting, 304, 306–307, 439t, 440
  deficiency, 439t
  in γ-carboxyglutamate formation, 306–307, 306f
  structure of, **306,** 439f
Vitamins
  B, 438, 438t
  coenzyme, 438–440, 438t
  defined, 438
  evolution of, 438–439
  noncoenzyme, 439, 439t
  in reducing homocysteine levels, 726
  roles of, 20
VJ recombination, 991–992, 991f
$V_{\max}$, 230
Voltage-gated ion channels, 383–384, 384f
von Gierke disease, 637, 637t, 638
von Gierke, Edgar, 637
VR1 (capsaicin receptor), 977–978, 977f, 978f
v-SNARE, 919
Wald, George, 971
Wang, James, 836
Warburg, Otto, 474
Warburg effect, 474, 554
Warfarin, 307
Water
  attack, facilitating, 277, 277f
  concentration of, 14
  dissociation of, equilibrium constant, 13–14
  as highly cohesive, 9
  hydrogen bonds, 9, 10
  as polar molecule, 8–9
  properties of, 8–9
Water-oxidizing complex (WOC), 574, 574f
Watson, James, 17, 109, 827
Watson-Crick base pairs. *See* Bases/base pairs
Watson-Crick DNA model, 110, 110f, 111, 116
Wernicke-Korsakoff syndrome, 821
Wernicke's encephalopathy, 515
Western blotting, 83–84, 83f, 138
White adipose tissue (WAT), 555
Whooping cough, 417–418
Wiley, Don, 996
Wilkins, Maurice, 109
Windows, 355
Withering, William, 373
Wobble hypothesis, 897, 897t
Wyman, Jeffries, 198

Xanthine oxidase, 761
Xanthomas, 785
Xanthylate (XMP), 752
Xenobiotic compounds, 1037
Xeroderma pigmentosum, 851
X-ray crystallography
  defined, 95
  diffraction patterns, 96, 96f
  electron-density map, 96–97, 96f, 97f
  experiment, 95, 95f
  reflections, 96
  resolution, 96, 97f
  synchrotron radiation and, 95

Yamanaka, Shinya, 947
Yanofsky, Charles, 935
Yeast artificial chromosomes (YACs), 147, 147f
Yeast chromosomes, 941, 942t

Z line, 1020, 1020f, 1021
Z scheme of photosynthesis, 577, 577f
Zantac (ranitidine), 1049, **1049**
ZAP-70, 999–1000
Zellweger syndrome, 657
Z-form DNA, 112, 112f, 112t
Zinc
  activation of water molecule, 265–267, 266f
  in biological systems, 265
  in carbonic anhydrase, 265, 265f
Zinc-finger domains, 945–946, 946f
Zinc-finger nucleases (ZFNs), 161–162, 162f
Zonal centrifugation, 77, 77f
Zwitterions, 29, 30f
Zymogens, 301–303, 301f
  cascade of activations, 303–304, 303f
  conversion into proteases, 302
  defined, 286
  developmental process control by, 299
  proteolytic activation, 301, 301f
  secretion of, 300, 300f

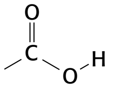
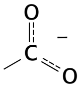
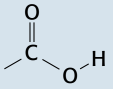
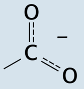
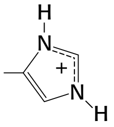
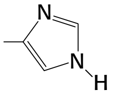
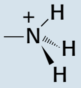
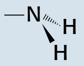
ACIDITY CONSTANTS

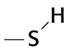
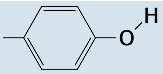
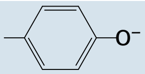
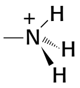
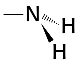
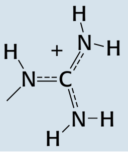
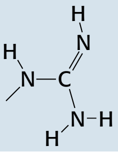
pK_a values of some acids

Acid	pK' (at 25°C)	Acid	pK' (at 25°C)
Acetic acid	4.76	Malic acid, pK_1	3.40
Acetoacetic acid	3.58	pK_2	5.11
Ammonium ion	9.25	Phenol	9.89
Ascorbic acid, pK_1	4.10	Phosphoric acid, pK_1	2.12
pK_2	11.79	pK_2	7.21
Benzoic acid	4.20	pK_3	12.67
n-Butyric acid	4.81	Pyridinium ion	5.25
Cacodylic acid	6.19	Pyrophosphoric acid, pK_1	0.85
Citric acid, pK_1	3.14	pK_2	1.49
pK_2	4.77	pK_3	5.77
pK_3	6.39	pK_4	8.22
Ethylammonium ion	10.81	Succinic acid, pK_1	4.21
Formic acid	3.75	pK_2	5.64
Glycine, pK_1	2.35	Trimethylammonium ion	9.79
pK_2	9.78	Tris (hydroxymethyl) aminomethane	8.08
Imidazolium ion	6.95	Water*	15.74
Lactic acid	3.86		
Fumaric acid, pK_1	3.03		
pK_2	4.44		

*$[H^+][OH^-] = 10^{-14}$; $[H_2O] = 55.5$ M.

Typical pK_a values of ionizable groups in proteins

Group	Acid	⇌	Base	Typical pK_a
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0

Group	Acid	⇌	Base	Typical pK_a
Cysteine		⇌	$-S^-$	8.3
Tyrosine		⇌		10.4
Lysine		⇌		10.0
Arginine		⇌		12.5

Note: pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group.

STANDARD BOND LENGTHS

Bond	Structure	Length (Å)
C—H	R_2CH_2	1.07
	Aromatic	1.08
	RCH_3	1.10
C—C	Hydrocarbon	1.54
	Aromatic	1.40
C═C	Ethylene	1.33
C≡C	Acetylene	1.20
C—N	RNH_2	1.47
	O═C—N	1.34
C—O	Alcohol	1.43
	Ester	1.36
C═O	Aldehyde	1.22
	Amide	1.24
C—S	R_2S	1.82
N—H	Amide	0.99
O—H	Alcohol	0.97
O—O	O_2	1.21
P—O	Ester	1.56
S—H	Thiol	1.33
S—S	Disulfide	2.05

ISBN-13: 978-1-4641-2610-9
ISBN-10: 1-4641-2610-0
90000
9 781464 126109